DNA Minor Groove Modifications: Synthesis and Application of 3-deaza-3-substituted-2'-deoxyadenosine Analogues

Author: Kerry Jane Salandria

Persistent link: http://hdl.handle.net/2345/2754

This work is posted on eScholarship@BC,
Boston College University Libraries.

Boston College Electronic Thesis or Dissertation, 2011

Copyright is held by the author, with all rights reserved, unless otherwise noted.

Boston College

The Graduate School of Arts and Sciences

Department of Chemistry

DNA MINOR GROOVE MODIFICATIONS:

SYNTHESIS AND APPLICATION OF 3-DEAZA-3-MODIFIED-2′-DEOXYADENOSINE ANALOGUES

A Dissertation

by

KERRY J. SALANDRIA

Submitted in partial fulfillment of the requirements

for the degree of

Doctorate of Philosophy

April 30, 2011

© by KERRY J. SALANDRIA

2011

For My Parents and Dan

Abstract

Nucleic acids are fundamental biomolecules responsible for all activities of a living cell. DNA serves as an instruction manual to the cell, containing blueprints and directions for all cellular processes, while RNA serves to carry out the messages held within DNA. Research into the structure, stability, and function of nucleic acids has revealed much about the origin and evolution of life.

The ultimate goal of this work is to understand how molecules bind and associate within the minor groove of double stranded, helical DNA. A series of 2′-deoxyadenosine analogues are modified at the three position by replacing the N^3-nitrogen with carbon. Substitution at this position is designed to emulate the effects of removing hydrogen bond acceptors, introducing steric bulk, and tethering functional groups of interest into the minor groove. These functional groups mimic small molecules that have been shown to bind within the minor groove of A-T rich sequences as well as serve as a platform for further substitution by fluorescent tags.

The synthetic effort needed to obtain purine nucleosides containing each of these modifications was non-trivial. New methodologies unveiled directing and protecting strategies towards the desired isomer of these modified nucleosides in higher yields than those previously deemed acceptable. Application of these modified nucleosides into duplex DNA reveals thermodynamic parameters for how small molecules bind to the minor groove and the effects of introducing biomarkers into an unprecedented region of DNA.

Table of Contents

Abbreviations and Acronyms

A adenosine

Ac acetyl

Ad adenine

AIBN azobis(isobutyronitrile)

Bn benzyl

Boc *tert*-butyloxycarbonyl

Bz benzoyl

calcd calculated

CD circular dichromism

d day(s)

dA 2′-deoxyadenosine

damc$_3$A 3-deaza-3-aminomethyl-2′-deoxyadenosine

DABCO 1,4-diazabicyclo[2.2.2]octane

DAP 2,6-diaminopurine

DBN 1,5-diazabicyclo[4.3.0]non-5-ene

DBU 1,8-diazabicyclo[5.4.0]undec-7-ene

dC 2′-deoxycytidine

dc$_3$A 3-deaza-2′-deoxyadenosine

DCM dichloromethane

dG 2′-deoxyguanosine

dhmc$_3$A 3-deaza-3-hydroxymethyl-2′-deoxyadenosine

DIPEA diisopropylethylamine

DMAP 4-(dimethylamino)pyridine

DMF *N*,*N*-dimethylformamide

dmc$_3$A 3-deaza-3-methyl-2′-deoxyadenosine

DMT 4,4′-dimethoxytrityl

DMSO dimethylsulfoxide

DNA β-D-2′-deoxyribonucleic acid

dtmc$_3$A 3-deaza-3-thiomethyl-2′-deoxyadenosine

Et$_2$O diethyl ether

EtOAc ethyl acetate

equiv equivalents

G guanosine

Gn guanine

h hour(s)

Hex hexanes

HPLC high performance liquid chromatography

HRMS high-resolution mass spectrum

M$_4$SA 2,2,3,3-tetramethylsuccinic anhydride

M$_4$SI 2,2,3,3-tetramethylsuccinimide

M molar

min minute(s)

NMR nuclear magnetic resonance

NBS *N*-bromosuccinimide

PAGE polyacrylamide gel electrophoresis

Pht phthaloyl

pyr pyridine

*R*f retention factor

rt room temperature

T thymidine

TBAF *n*-tetrabutylammonium fluoride

TBS, TBDMS *tert*-butyldimethylsilyl

TCA trichloroacetic acid

TEA triethylamine

Tf triflate (trifluoromethanesulfonyl)

TFA trifluoroacetic acid

TLC thin layer chromatography

TMS trimethylsilyl, also tetramethylsilane

Tol *para*-toluoyl

Tr trityl (triphenylmethyl)

Ts tosyl

RNA β-D-ribonucleic acid

U uridine

UV Ultraviolet

Chapter 1: Introduction

1.1 Early Nucleic Acids Research

Nucleic acids are the foundation and basis of all life forms. The oldest fragment of DNA dates back over 400 million years ago and is from the time when bacteria ruled Earth.[1] Today, nucleic acids research has become prominent in a diverse set of fields. Its applications in medicine, agriculture, and nanotechnology have demonstrated its widespread function to benefit society. Its role and the molecules associated with DNA allow us to understand life on a level that was not possible 150 years ago. Meischer was the first to discover DNA in 1869 by exposing the nucleus of a cell to alkaline and acid conditions, leading to the formation of a precipitate.[2] At the time, it was referred to as nuclein, since it is mainly stored in the nucleus of the cell. Over the next 30 years, important discoveries regarding the components of the newly coined "nucleic acids" and the enzymes associated with them led to a Nobel Prize in Physiology or Medicine to Albrecht Kossel in 1910. Kossel is credited with deciphering the genetic code of DNA into its four monomers: 2′-deoxyadenosine, 2′-deoxythymidine, 2′-deoxyguanosine, and 2′-deoxycytidine.[3] It would later be revealed that these monomers form long chains through phosphodiester linkages (**Figure 1**). The structure and role of DNA within the cell was further elucidated in the following years. Proof that DNA is hereditary material came in 1944. Avery, MacLeod and McCarty extracted viral DNA from pathogenic *Streptococcus pneumonia* and transformed it into a non-viral strain of the same bacteria.[4] The non-viral bacteria became virulent after transformation, demonstrating that DNA is

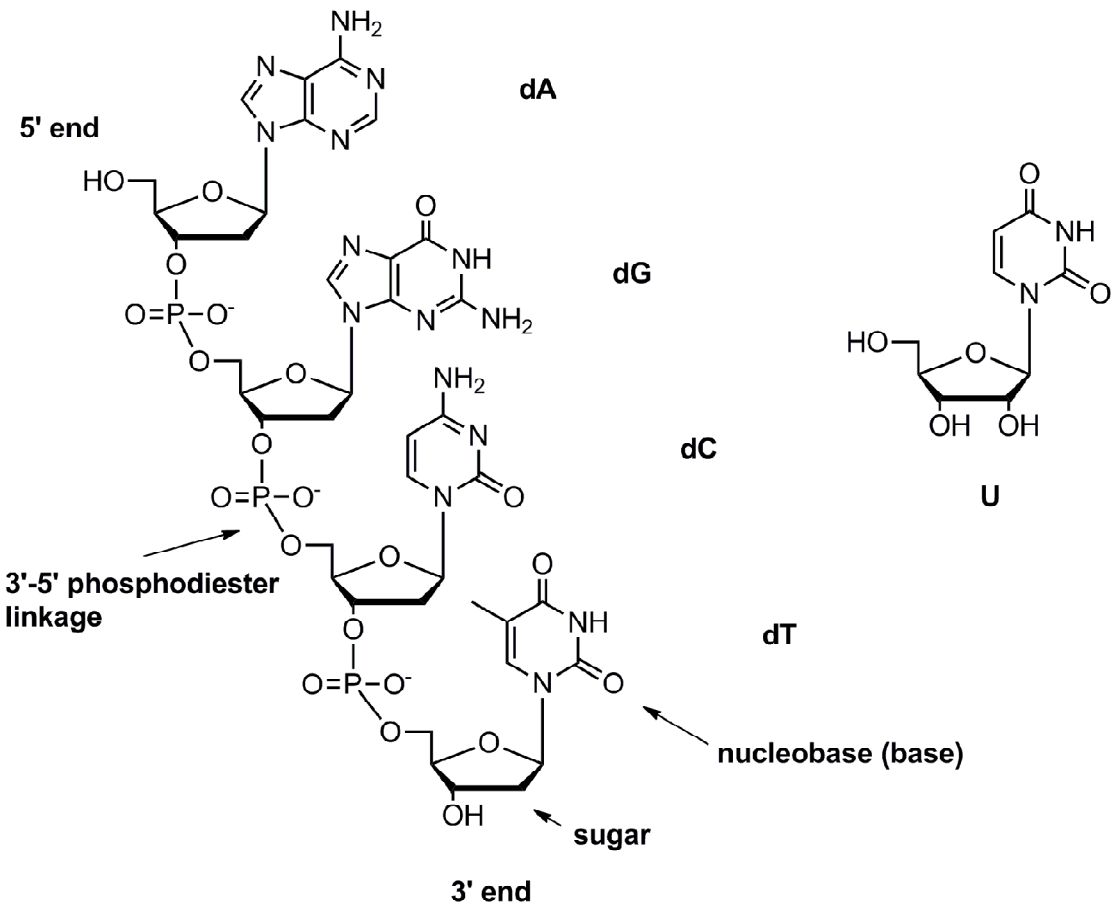


Figure 1. A structure of a polynucleotide containing the four residues (dA, dG, dC, and dT) as they would appear in DNA. In, RNA, the 2'-hydroxyl is present on all sugars and U pairs with A instead of dT.

the material responsible for genetic differences within the same organism, not proteins or other cellular components.

The discovery of the double-helical structure of DNA is widely credited to Watson and Crick in 1953. This discovery would not be possible without previous results by Astbery, Bell, Chargaff, and Wilkins. In 1938, Astbery and his graduate student,

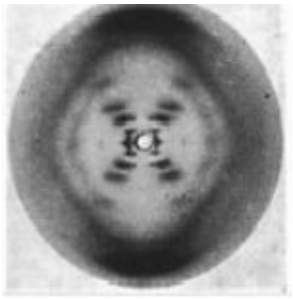


Figure 2. X-ray diffraction pattern of DNA obtained by Wilkins and Franklin.[7]

Florence Bell, showed that the bases in a DNA molecule are stacked one above the other and lie perpendicular to the long axis of the molecule "as if they were a pile of pennies".[5] "Chargaff's Rule" states that the number of pyrimidines is equal to the number of purines within DNA,[6] suggesting some type of pairing between the two heterocycles. In 1953, Franklin and Wilkins were the first to obtain an X-ray crystal structure of DNA and publish it in *Nature* (**Figure 2**).[7] The analysis demonstrated that DNA has a regularly repeating structural unit every 3.4 Å, with a complete turn of the axis occurring every 34 Å, or every 10 bases. Without knowledge of this crystal structure, Watson and Crick

wrote to *Nature* describing a model structure for DNA that is still used today.[8] They described the structure as a double helix that is held together by a pairing system. There are two base pairs in DNA: dA pairs with dT and dG pairs with dC (**Figure 3**). Their proposed secondary structure directly supported a copying mechanism of the primary structure and a means of passing genetic information from generation to generation. This discovery transformed nucleic acid research forever and has allowed scientists to discover key functions of DNA as genetic material.

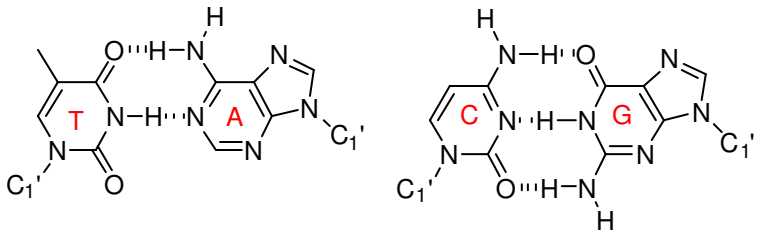


Figure 3. The two Watson-Crick base pairs present in DNA. In RNA, U replaces T.

1.2 Nucleic Acid Structure

The discovery of the double helix by Watson and Crick opened a door to researchers studying nucleic acids. The secondary structure allowed researchers to go in two opposing directions: microscopic, examining the molecular structure of each monomer and macroscopic, examining how nucleic acids behave as macromolecules. While many foundations been laid, how the microscopic structure affects the macroscopic structure is still a current focus of nucleic acids research and is a main theme of the work described in this thesis.

From a molecular point of view, nucleic acids are composed of building blocks referred to as nucleotides (**Figure 1**). Each nucleotide has three key features: a nucleobase, a sugar, and a phosphate group. If the pyrimidine or purine is free, it is referred to as a nucleobase. Once coupled to a sugar, they are referred to as nucleosides (**Table 1**).

Table 1. Nucleic Acids Nomenclature

Nucleobase	Nucleoside	Example of Nucleotide
Guanine	Guanosine	Guanosine-5′-monophosphate
Cytosine	Cytidine	Cytidine-3′-monophosphate
Adenine	Adenosine	Adenosine-3′,5′-bisphosphate
Thymine	Thymidine	Thymidine-5′-triphosphate
Uracil	Uridine	Uridine-2′-monophosphate

In nature, guanine, cytosine, and adenine are coupled to both ribose and 2-deoxyribose sugar moieties. The nomenclature in **Table 1** is for ribonucleosides. If these bases are coupled to 2-deoxyribose, the names of these nucleosides are simply prefixed with 2′-

deoxy. It is important to note that in the cell, thymine is coupled with 2-deoxyribose to produce thymidine (dT) and uracil is coupled with ribose to produce uridine (U). The form of sugar is assumed within the nomenclature for these two nucleosides.

Nucleotides are linked by a single phosphate group bridging the 3'-hydroxyl of one monomer to the 5'-hydroxyl of the next monomer by a phosphodiester bond, as shown in Figure 1. These linkages impart stability and longevity to nucleic acids and are a key reason why they have evolved to store and pass genetic information from one cell to another over multiple generations. Phosphodiester linkages are negatively charged at neutral pH which prevents them from crossing cellular membranes, holding them within the nucleus of the cell. Their inherent negative charge offers kinetic stability to nucleophilic attack, providing resistance to hydrolysis, which explains how these molecules have lasted millions of years in an aqueous environment with little damage to their integrity.[9]

Multiple nucleotides linked together form single strands of ribonucleic acids (RNA) or deoxyribonucleic acids (DNA). Nucleic acids are capable of forming a diverse set of secondary structures (**Figure 4**).[10] Depending on the sequence, they can fold back and base pair with themselves to form β-hairpins. The sequence of this type of structure is typically palindromic, which imparts its self-complementary character. The β-turn of the hairpin requires a minimum of 3 nucleotides to form a stable structure. Non-palindrome sequences can still form hairpins, but these structures are naturally less stable due to bulges of unpaired nucleotides. Non-palindrome sequences prefer to pair with a

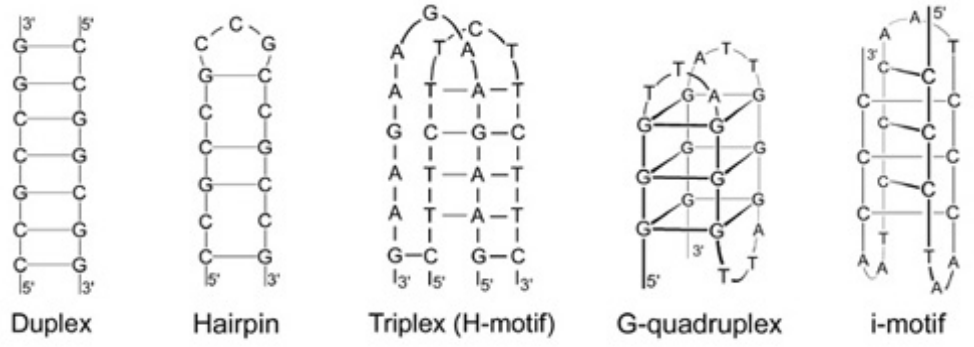


Figure 4. Various secondary structures of DNA.[10]

second strand of DNA/RNA containing the complement to form double stranded DNA/RNA. The two complementary strands run antiparallel (*i.e.* 5'→3' complements with 3'→5' strands) with the phosphodiester linked sugars forming an external, anionic backbone. The nucleobases lie perpendicular to the backbone, stacked upon one another to form a hydrophobic core. The formation of the helix results in two grooves. These two grooves are commonly referred to as the major and minor grooves (**Figure 5a**). Imagine the phosphate backbones of each strand lying on the perimeter of a circle, with the base pairs cutting across the diameter of the circle. The grooves are defined as the distance on the circle between the phosphates on the backbone, with the larger being the major groove and the smaller the minor groove. The base pairs are arranged in such a way that certain functional groups are characteristic to each groove (**Figure 5b**). For example, the methyl group on thymine is located in the major groove, while the amine on guanosine is pointed towards the minor groove. Differing sequences of DNA have different patterns of

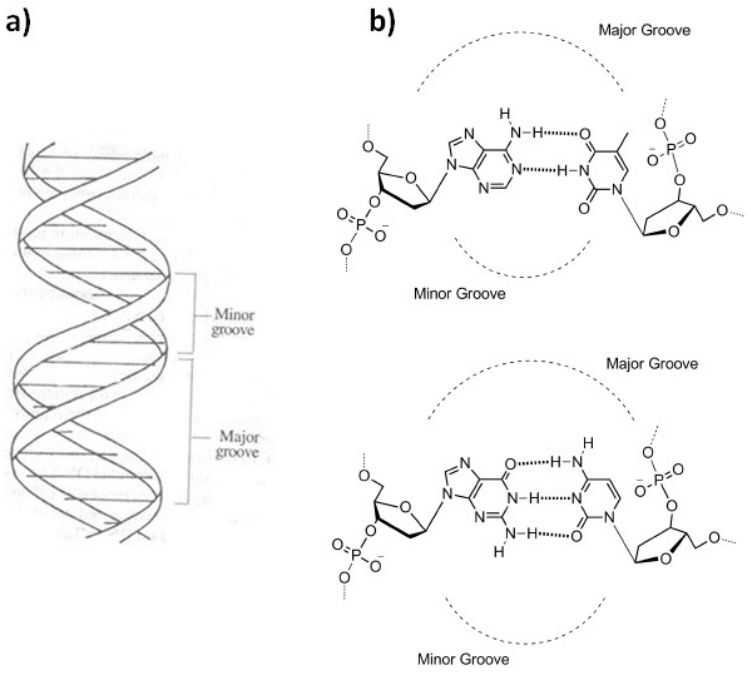


Figure 5. a) Major and minor grooves of a double helix. **b)** Major and minor grooves of A-T (above) and G-C (below) base pairs and the functional groups that lie in each.

functionality within their grooves that biomolecules, small molecules, and cations use as a fingerprint to bind to DNA with sequence specificity.

The helix itself can take on different macroscopic structures depending on the conformation of its monomers. The three predominant forms of nucleic acids are A-, B-, and Z- forms (**Figure 6**). Other forms of DNA have been described within the literature, but exist under atypical conditions.[11] There are many unique characteristics to each form of helix which arise from various components within the structure of the monomer. Before discussing the structural aspects of A-, B-, and Z- form nucleic acids, it is important to review these components.

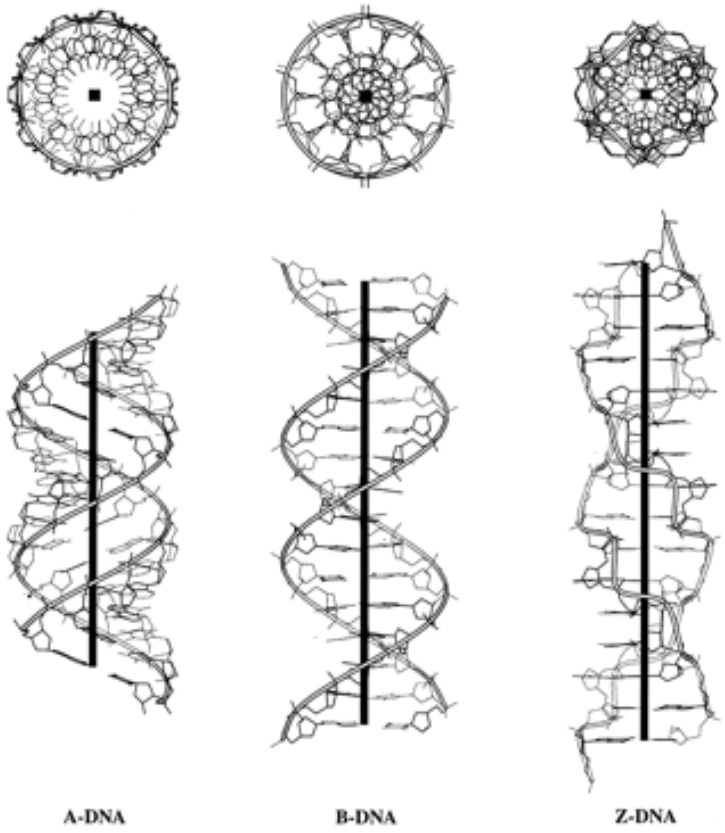


Figure 6. A-, B-, and Z- duplex structures. Above: top-down view of the helix. Below: Side view of the helix.[12]

The major component affecting the conformation of the helix is the sugar pucker of each monomer. The five atoms within a furanose ring rarely exist in the same plane; one or two atoms typically lie out of the plane. D-Ribose and 2-deoxy-D-ribose predominantly exist in two types of sugar pucker, C3′-*endo* and C2′-*endo*, which are responsible for A- and B-form nucleic acids, respectively (**Figure 7a**). Less commonly observed is the C2′-*exo* sugar pucker which is found in Z-form DNA. Another component of each monomer's conformation is the orientation of the nucleobase with respect to the

sugar. This orientation changes by rotation about the glycosidic bond. Nucleosides prefer to exist in *anti*-orientation, where the Watson-Crick hydrogen-bonding face of the nucleobase is pointed away from the sugar. In *syn*-orientation, the hydrogen-bonding face is pointed over the sugar. This orientation is typically disfavored due to sterics of the nucleobase with the 5'-carbon on the sugar (**Figure 7b**). The sugar pucker and orientation of the nucleobase have an additive affect on each base pair, leading to a global effect on the helix.

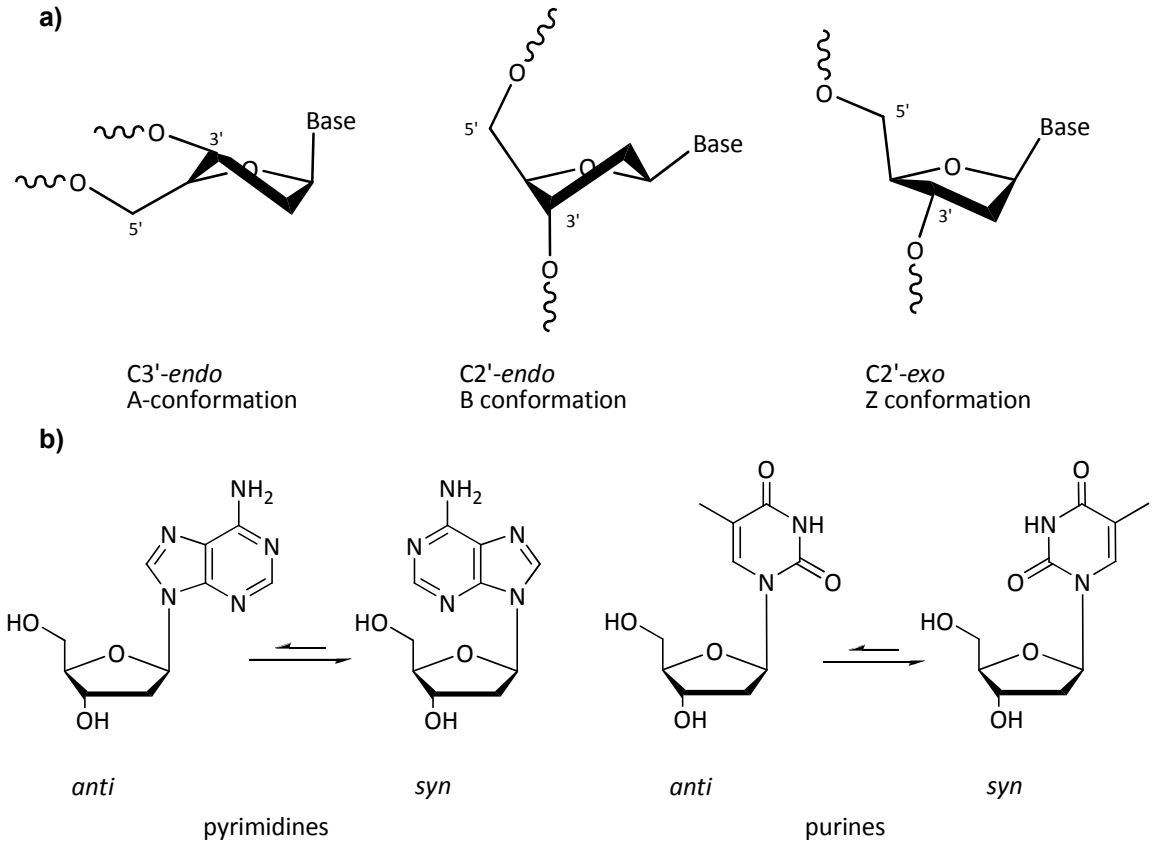


Figure 7. a) Common sugar puckers of 2-deoxy-D-ribose and the resulting helix conformation. **b)** Illustrations of *anti* and *syn* conformations of nucleosides.

A-, B- and Z- form helices are easily identifiable from one another by examining their crystal structures. The diameter and depths of the grooves in each helix are noticeably different between forms. Perhaps more subtle is the positioning of the base pairs. There are many parameters available to describe the location of the base pairs within the helix (**Figure 8**). Most important for describing the base pairs within each conformation of helical DNA are inclination, propeller twist, and twist. Inclination refers to a degree of rotation about the vertical axis of the base pairs, where 0^o infers the base pairs are perpendicular to the vertical axis of the helix. The heterocycles in Watson-Crick

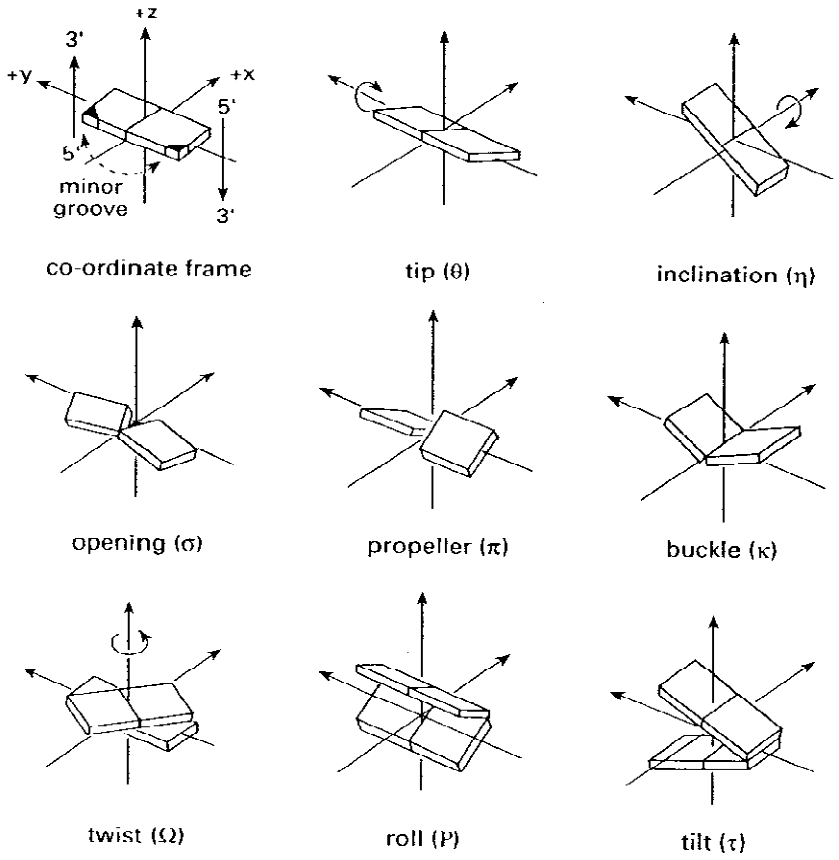


Figure 8. Structural parameters used to describe base pair positioning. Figure adopted from *Principles of Nucleic Acid Structure* by Wolfram Saenger

base pairs are not coplanar and the degree to which the bases are rotated is referred to as propeller twist. This is not to be confused with twist, the degree of which each base pair is rotated from the previous base pair along the length of the helix. Other parameters include rise, the distance between each base pair along the helical axis, and pitch, the measured distance of one full turn of the helix. Each form of helix has "hallmark" characteristics that are discussed below.

A-form is a right-handed helix with the widest diameter and smallest rise per base pair of any form, both attributed to the sugar pucker. The C3′-*endo* sugar pucker places the 3′- and 5′- hydroxyls within 5.9 Å of one another, leading to its small rise and wide diameter. The large diameter places the base pairs off-center of the vertical axis, resulting in an exaggerated major and diminished minor grooves. The major groove is very deep, while the minor groove is extremely shallow. Base pairs in an A-form conformation are on a severe incline of $+19^{o}$. RNA-RNA or RNA-DNA helices appear to exist predominantly in A-form due to the presence of the 2′-hydroxyl on ribose. This hydroxyl places the sugar in a C3′-*endo* conformation, which allows intramolecular hydrogen bonding to the C4′-oxygen within the ring of the furanose. In order to observe A-form DNA-DNA helices, dehydrating conditions of high salt and duplex DNA are required. These are common crystallization conditions, so it remains unclear as to whether or not DNA-DNA helices exist in A-form under normal cellular conditions.

DNA-DNA helices primarily exist as B-form structures within the cell. B-form DNA is more narrow and elongated than A-form due to C2′-*endo* sugar puckering

placing the backbone hydroxyls 7.0 Å apart. This form of helical DNA places the base pairs on the center of the vertical axis, resulting in more equally sized major and minor grooves. The base pairs are nearly perpendicular to the vertical axis of the helix with a small inclination of -1.2^{o}. B-form DNA is less distorted and more symmetrical than both A- or Z-form helices.

Z-form helices are plausibly the most distorted form of DNA within cellular conditions. Its role in nature has been difficult to elucidate, but is thought to relieve supercoiling of DNA while transcription occurs.[13, 14] This helix is left-handed in nature and generally disfavored. Alternating poly(dGC) sequences in the presence of high salt can encourage B- to Z-form transitions. Interestingly, the sugar pucker in these alternating poly(dGC) sequences is C2′-*endo* for dC residues and C2′-*exo* for dG residues. Z-form DNA is the only naturally occurring helix with *syn* conformation of dG residues. These features cause the repeating unit of Z-form to be 2 base pairs, as opposed to a repeating unit of 1 base pair for A- and B-forms. It has the narrowest diameter of all 3 forms and a severe degree of rotation per base pair (60^{o} versus ~35^{o} for A- and B-forms). The three forms and their key structural features are highlighted in **Table 2**.[15, 16]

CD spectroscopy is a powerful tool in elucidating the overall structure of helical DNA and RNA. In this type of spectroscopy, the sample is exposed to circularly polarized light of varying wavelengths. The circularly polarized light rotates in either a clockwise or counter-clockwise direction. Different wavelengths are absorbed by the sample, and this absorption also differs depending on the direction of the circularly

Table 2. Structural Parameters of A-, B-, Z-form DNA Configurations.

Parameter	A-form	B-form	Z-form
Helix Sense	Right-handed	Right-handed	Left-handed
Repeating Unit	1 bp	1 bp	2 bp
Rotation/bp	32.7°	35.9 °	60°/2
bp/turn	11	10.5	12
Inclination	19°	-1.2°	-9°
Rise	2.3 Å	3.32 Å	3.8 Å
Pitch	28.2 Å	33.2 Å	45.6 Å
Propeller Twist	18°	16°	0°
Glycosyl Angle	*anti*	*anti*	C: *anti* G: *syn*
Sugar Pucker	C3′-*endo*	C2′-endo	C: C2′-*endo* G: C2′-*exo*
Diameter	23 Å	20 Å	18 Å

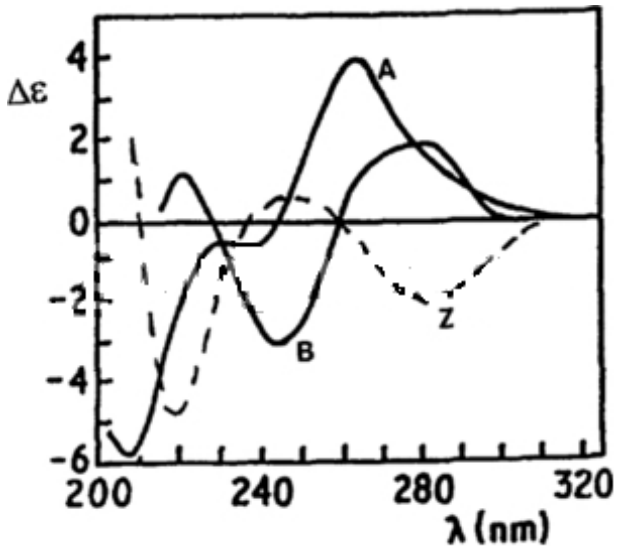


Figure 9. General CD Spectra of A-, B-, and Z-form DNA configurations. Image taken from Springer Images.

polarized light. The difference in absorbance between left- and right- polarized light is plotted as a function of wavelength to give a unique spectrum specific to the overall structure of DNA or RNA, whether it be A-, B-, or Z- form (**Figure 9**).[17] Spectra of the same sample can differ as a function of duplex and salt concentrations. The general pattern of ellipticity indicates the form of duplex (A-, B-, or Z-). When salt and duplex concentrations are held constant, differences in intensity can offer some insight to perturbations of the structure. Base-stacking, handedness of the helix, and overall disturbances of the helix can be observed in the CD spectra.

1.3 Nucleic Acid Thermodynamic Stabilization

The different structures of nucleic acids are a direct result of the environment of the helix. Increases in salt concentration have demonstrated B- to Z- transitions[18] and additional hydrogen bonding has shown how identical sequences of RNA or DNA preferentially exist in A-[19] or B-form[20], respectively. These and other factors contributing to the stability of helical DNA are partly responsible for the molecules' longevity and ability to pass genetic information.

The thermodynamic stability of duplex nucleic acids can be measured using thermal denaturation curves (**Figure 10**). In these types of curves, hybridized DNA (or RNA) at low temperature is slowly warmed to higher temperatures while monitoring the

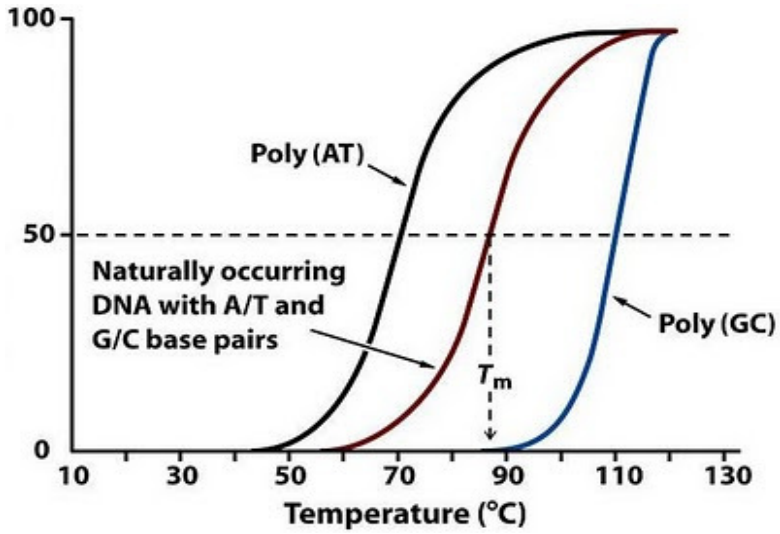


Figure 10. Thermal Denaturation Curves of DNA depicting the relationship between sequence and stability. Source: Wikipedia.

absorbance (λ_{max} 260 nm). As the duplex is heated, it unwinds to single strands. Single stranded DNA has a greater absorbance than duplex DNA due to a hyperchromic effect. This effect is the result of multiple aromatic residues stacked upon one another within the center of the helix, significantly quenching the total absorbed light per base pair. As the helix is denatured, these aromatic resides are more exposed and thus able to absorb more light. The temperature at 50% denaturation is considered the melting temperature, or T_m, and is reflective of the overall stability of duplex DNA.

Three factors that significantly contribute to nucleic acid stability are hydrogen bonding, base stacking, and phosphate stabilization. Hydrogen bonding is understood to occur specifically between each base pair, which lends to a direct mechanism for replication and transcription. Adenosine base pairs with thymidine or uridine through two hydrogen bonds: one occurs at the N^6-amine on A to the O^4-carbonyl on T/U and the second occurs within the heterocycles at the N^1 and N^3 positions of A and T/U, respectively. Sequences rich in G-C base pairs have been shown to be more stable than sequences rich in A-T base pairs because of the presence of a third hydrogen bond (**Figure 3**). A second indication that hydrogen bonding affects duplex stability is that longer sequences of polynucleotides are more stable than shorter sequences due to the presence of additional hydrogen bonds. An added effect of longer sequences that leads to a higher stability is base stacking.

At its core, DNA contains a series of aromatic residues stacked upon one another. Base-stacking, or π-π stacking, results from interactions of electrons in unhybridized p-

orbitals. These orbitals are able to interact in an expected manner on the plane of the molecule and also with the π-orbitals of base pairs above and below the plane of the base pair. This effect is not only a type of van der Waals intermolecular force, but also allows dipole-dipole interactions of DNA. These effects give a net effect that stabilizes the duplex and keeps the structure intact.

Hydrogen bonding and base stacking favor the formation of the duplex, but this favorability is matched with an opposing force. Each strand of DNA is negatively charged and bringing these backbones within proximity of one another results in negative charge-charge repulsion. The concentration of cations within the environment is directly proportional to the stability of the duplex. Increasing the concentration of salt within solution has been shown to give higher melting temperatures, with the effect leveling off at approximately 1 M.[21] These cations have been shown to closely associate with the negative backbone, as well as in regions where the two strands are within their closest proximities. Specifically, sequences containing A-tracts (4 or more consecutive A-residues) have a narrower diameter of the helix than random sequences in the 5' to 3' direction. These narrow regions have been shown to bind ammonium cations by ^{15}N-NMR studies.[22] If the A-tract is followed by a T-tract, the helix reaches a minimum at the ApT step (**Figure 11, left**). The narrowest region occurs at the fourth consecutive A-residue and continues until the A-tract ends (**Figure 11, middle**). If an A-tract is preceded by a T-tract, the widest region of the helix in this motif will be the in TpA step (**Figure 11, right**). Minimizing the diameter of the helix brings

Figure 11. Structural Effects of A-tract DNA and Localization of Ammonium Cations (circles). Left: The narrowest region occurs at the ApT step of sequences containing A_4T_4 tracts. Middle: The diameter of the helix reaches its minimum at the 4th consecutive A and continues until an ApX step is reached. Right: Sequences containing T_4A_4 tracts are narrowest at the outer regions. Image taken from Reference 22.

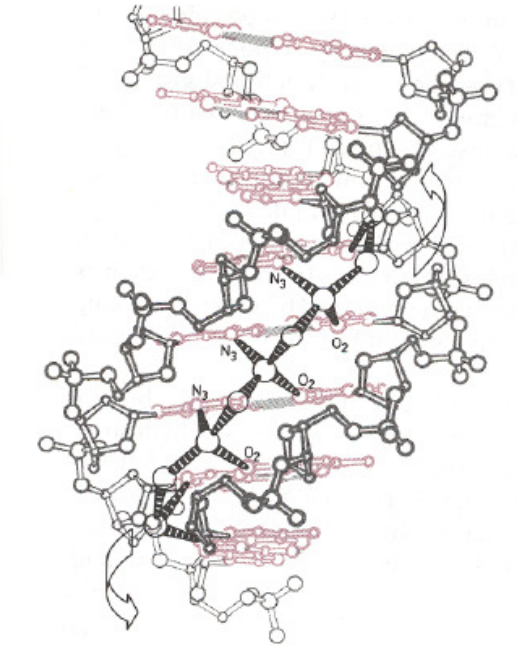


Figure 12. DNA Hydration in the Minor Groove occurs between the N^3 of dA and O^2 of dT, on the opposite strand one base pair displaced. A second layer of water bridges the first layer to offer more hydrogen bonding and structural reinforcement.

opposing backbones very closely together in the minor groove, forming negatively-charged electrostatic pockets which bind monovalent cations.

Neutral molecules other than ammonium have also been shown to bind to A-tract DNA and contribute to the overall stability of the duplex.[23-25] In particular, hydration within the minor groove of A-T rich DNA offers additional hydrogen bonding and structural reinforcement. The first insights into the significance and contribution of water binding to duplex DNA was in 1981. The crystal structure of 5'-d(CGCG AATT CGCG) revealed many interesting facets of B-form DNA.[26, 27] This structure had such an impact on nucleic acids research, that it is now commonly referred to as the Dickerson dodecamer. Importantly, a specific ordering of water molecules in the minor groove was shown (**Figure 12**). Water molecules hydrogen bond with the N^3 of adenine and O^2-carbonyl of thymine on the opposite strand, one base pair displaced to form a primarily shell of hydration. A second shell forms through water molecules bridging together the first shell. It is now understood that this hydration pattern can occur in any A-T rich sequence because the hydrogen bond acceptor on dA and dT are within the same proximity in the minor groove in B-form DNA. The N^2-amine on dG residues protrudes into the minor groove and sterically blocks hydration effects in G-C rich sequences of B-form DNA. A-form DNA does not demonstrate this effect since the minor groove is too shallow.[28]

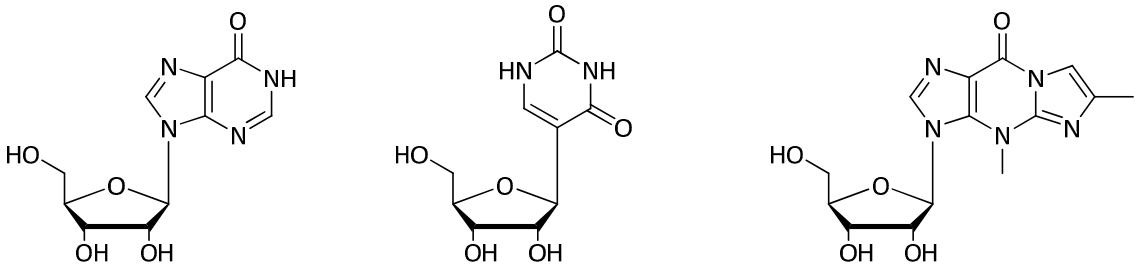
Minor groove hydration of A-T rich sequences in B-form DNA has been shown to contribute 0.7 kcal/mol per base pair.[29, 30] This value was obtained by incorporation of

modified dA and dT residues which delete the water's hydrogen bond acceptors at the N^3 of dA and O^2 of dT. Nucleoside analogues can be used to probe the structure and stability of duplex DNA and are an important contributor to understanding intermolecular forces and their role within the helix.

1.4 Nucleoside Analogues

Nucleosides are defined as a heterocycle bound to a ribose sugar via a β-glycosidic bond (*R* stereochemistry at the C1' of D-ribose). Adenosine, guanosine, cytidine, and thymidine exist in chromosomal DNA, but the cell generates multiple nucleoside analogues for different metabolic purposes (**Figure 13**).[32] Inosine, for example, is used in RNA. Inosine is capable of wobble base pairing to adenosine, thymidine, and cytidine. Incorporation of inosine into tRNA allows the molecule to contain the same anti-codon and pair with different codons, thus conserving the cells energy in the manufacturing of tRNAs.

Analogues can be synthesized *de novo* by the cell for purposes such as those discussed above. Synthetic chemistry has made a plethora of analogues accessible; some are inspired by native nucleosides while others demonstrate the creative freedoms chemists can take outside of physiological conditions. These analogues can be used



inosine pseudouridine quesuonine

Figure 13. Nucleoside Analogues used found in tRNA.

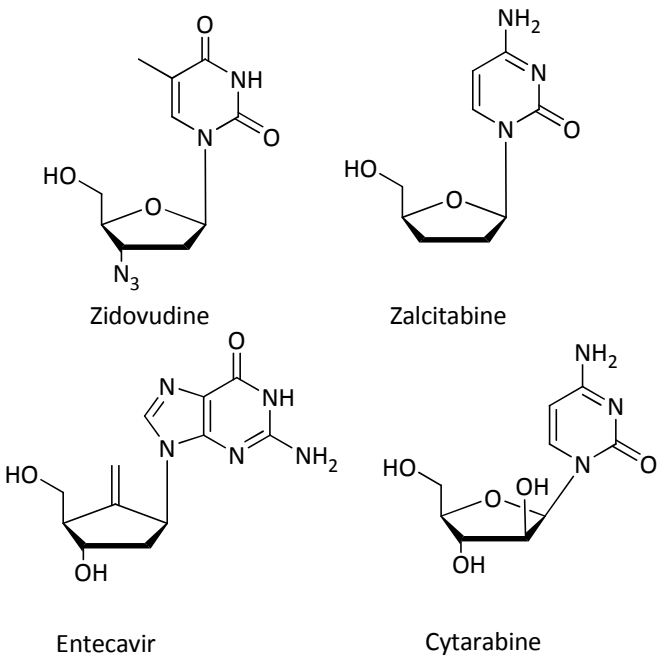


Figure 14. FDA- Approved Nucleoside Analogues

medicinally as antivirals or chemotherapeutics (**Figure 14**). Chain terminators, such as Zidovudine or Zalcitabine, lack the 3′-hydroxyl necessary for addition of sequential nucleosides during replication. Other sugar modifications include those seen in Entecavir and Cytarabine. Entecavir has been shown to block the active site of viral DNA polymerase. Cytaribine is used as a chemotherapeutic, but works in the same principal by blocking DNA and RNA polymerases in rapidly dividing cells.

In the past 40 years, many modified nucleoside analogues have been chemically synthesized for laboratory purposes such as fluorescence-mediated binding of molecules to nucleic acids, purification, and structure/stability studies. These modifications commonly occur on the backbone,[32] at the 5'[33] or 3'[34] terminus, or within the major

groove of nucleosides.[35] Fluorophores and reporter groups are typically attached to the C^5-methyl on thymine, the exocyclic amine of cytidine, or either terminus of polynucleotides, but they have also been shown to attach to amine and thiol modified phosphodiester backbones.[36] Stability and structural studies have shed light on the evolution of DNA and RNA as well as the limitations and tolerances of the helix within certain environments.

An elegant study by Barry Gold and coworkers used nucleoside analogues to demonstrate the importance of small molecule binding in the major groove of DNA and serves as important background for the work discussed in this thesis.[37] In this study, a range of functionality was synthetically introduced into the 7-position of guanosine and incorporated into self-complementary oligonucleotides to study their effect on the helix (**Figure 15**). The major findings of this study show the thermodynamic importance of the N^7-G in binding water and ammonium (**Table 3**).

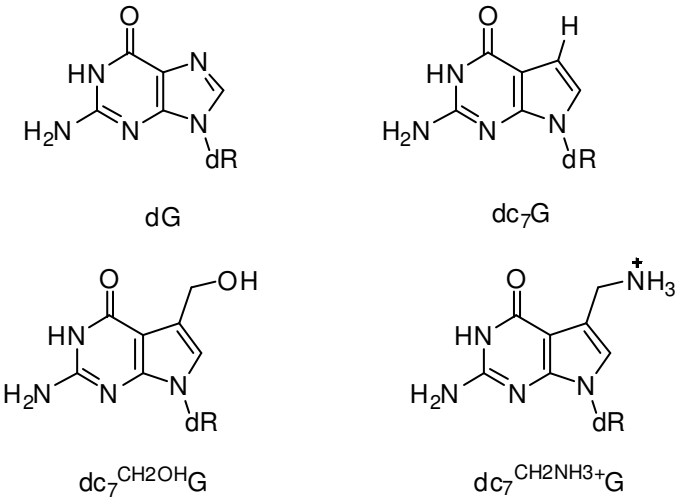


Figure 15. Analogues of dG used to probe hydration and ammonium binding in the major groove.

Table 3. Thermodynamic Effect of Various Substitutions in the Major Groove of Duplex DNA.

	Sequence	Tm (^{o}C)[†]	ΔH, kcal/mol	TΔS,[‡] kcal/mol	ΔG, kcal/mol
1	d(GAGAGCGCTCTC)	48.7	-78.2	-71.3	-6.9
2	d(GAGAc_7GCGCTCTC)	44.7	-56.3	-51.9	-4.4
3	d(GAGA$c_7^{CH2OH}G$CGCTCTC)	47.2	-54.5	-49.9	-4.6
4	d(GAGA$c_7^{CH2NH3+}G$CGCTCTC)	52.0	-92.9	-83.8	-9.1
5	d(CGCGTTTTCGCG)	68.4	-31.0	-26.6	-4.4
6	d(CGC$c_7^{CH2NH3+}G$TTTTC$c_7^{CH2NH3+}G$CG)	65.6	-27.0	23.4	-3.4

[†]Conditions: 10 mM Phosphate Buffer (pH 7.0), 10 mM NaCl, 7 uM duplex. [‡]Determined at 20 ^{o}C.

Upon a single deletion of the N^7-G (**entry 1**), the melting temperature decreased by 4 ^{o}C, corresponding to 2.5 kcal reduction in the favorability of ΔG. Tethering a hydroxyl mimic (**entry 2**) to this position increased the melting temperature relative to c_7G by 2.5 ^{o}C, but had little effect on ΔG. The most interesting effect resulted from incorporation of a tethered ammonium into this same position (**entry 4**). This duplex is more stable than those containing native dG as shown by a 3 ^{o}C increase in the T_m, which correlates to a 2.2 kcal of stability in ΔG. However, incorporation of multiple ammonium substituted analogues into a self-complementary sequence results in a decreased melting temperature, due to the close proximity of each ammonium (**entries 5 and 6**).

The stability effects observed through these substitutions also translated into alterations within the structure of the helix. CD spectroscopy of all the substitutions showed a general B-form helix (as described in section 1.2). Interestingly, the negative intensities of each duplex reflect changes in ΔH (**Figure 16**). The differences in each spectrum are understood to correlate to changes in base stacking effects.

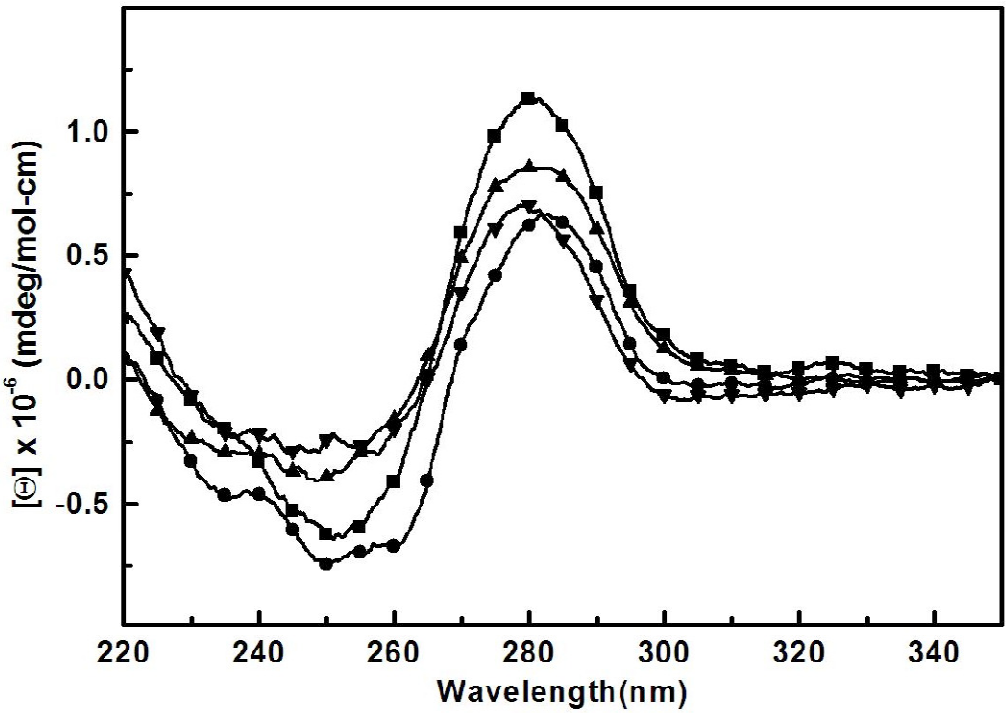


Figure 16. CD Spectral Effects of Various Substitutions at the 7-position of G. Entry 1: Squares; Entry 2: Triangles, Entry 3: Upside Down Triangles; Entry 4: Circles (see Table 3).[37]

1.4 3-deazapurines: Background and Applications

A particular class of modified nucleosides used throughout the studies described in this thesis are 3-deazapurine nucleosides. IUPAC nomenclature for this type of heterocycle is referred to as imidazo[4.5-*c*]pyridine, but researchers in this field have adopted their own purine numbering system that differs from traditional IUPAC numbering (**Figure 17**). Both names are used interchangeably, sometimes by the same author in the same manuscript; they will be referred to here-on by the purine numbering system. It should be understood that 3-deaza-2′-deoxyadenosine (dc_3A) refers to replacement of the N^3 by C-H and any other modifications at this position will be referred to as 3-deaza-3-substituted-2′-deoxyadenosine (dxc_3A) (**Figure 18**).

These modified purine systems are found in nature, particularly 3-deazaadenosine, both in green leaves of spinach and the cutaneous glands and liver of sharks and amphibians.[38] While little is known about its biological purpose in these organisms, it exhibits anti-inflammatory and anti-viral activity. The biological purposes of these molecules should not be undermined, but 3-deazapurines can be used to reveal important structural and stabilizing features of the minor groove of DNA. Removal of the nitrogen allows tethering of small molecules into the minor groove without creating a positively charged heterocycle. These modified nucleosides can be used to gain further insight into the minor groove of DNA and is the purpose of this work.

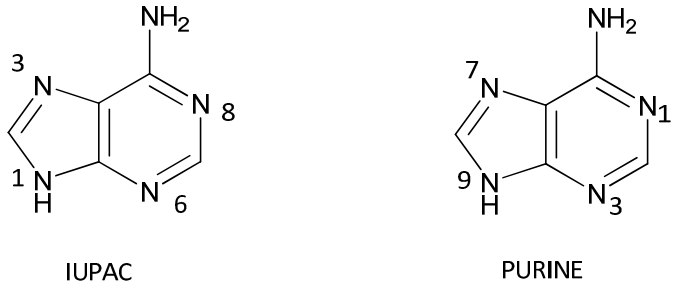


Figure 17. Numbering Systems for Purines. Left: Traditional IUPAC System. Right: System Adopted by Nucleic Acid Researchers.

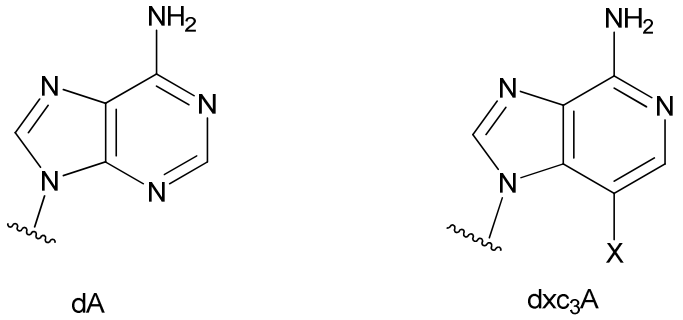


Figure 18. General Structure of 3-deaza-3-substituted-2'-deoxyadenosine.

References

1) Ganguly, M., Wang, R., Marky, L. A., Gold, B. *J. Am. Chem. Soc.*, **2009**, *131 (34),* 12068–12069.
2) Ralf Dahm. *Dev. Bio*. **2005**, *278*, 274–288.
3) a) Kossel, A. *Z. Physiol. Chem.* **1879**, *3*, 284-291. b) Kossel, A. *DuBois-Reymond's Arch.* **1891**, *181*, 181-186.
4) Avery, O.T., MacLeod, C.M., McCarty, M.*, J. Exp. Med*. **1944**, *79*, 137– 158.
5) Olby, Robert. *The Path to the Double Helix : The Discovery of DNA*. Seattle: University of Washing Press, 1974; p 65.
6) Chargaff E, Lipshitz R, Green C, Hodes ME. *J Biol Chem.* **1951**, *192* (1), 223–230.
7) Franklin, R. E.; Gosling, R. G. *Nature*, **1953**, *172* (4369), 156–157.
8) Watson, J. D. Crick, F. H. C. *Nature*, **1953**, *171*, 737-738.
9) Westheimer, F. *Science,* **1987**, *235 (*4793*),* 1173-1178.
10) Jaumot J, Eritja R, Navea S, Gargallo R. *Anal Chim Acta.* **2009**, *642(1-2),* 117-126.
11) Neidle, S. *Oxford Handbook of Nucleic Acid Structure*, Oxford Press, 1999: 1-38.
12) Lu, X. J. and Olson, W. K. *Nucl. Acids Res.* **2003**, *31(17)*, 5108-5121.
13) Ha, S. C., Lowenhaupt, K.; Rich, A., Kim, Y. G., Kim, K. K. *Nature*, **2005**, *437(7062)*, 1183–1186.
14) Rich, A., Zhang, S. *Nature Review Genetics,* **2003**, *4(7)*, 566–572.

15) Rich, A. Norheim, A., Wang, A. H. J. *Annual Review of Biochemistry*, **1984**, *53 (1)*, 791-846.

16) Ho, P. S. *Proc. Natl. Acad. Sci, USA,* **1994**, *91 (20),* 9549-9553.

17) Figure adopted from: Springer Images, www.springerimages.com, © 2011 Humana Press, Totowa, NJ.

18) Zhang, H., Yu, H., Ren, J., Qu, X. *Biophysical Journal,* **2006**, *90(9)*, 3203–3207.

19) Broyde, S. and Hingerty, B. *Nucl. Acids Res*, **1978**, *5(8),* 2729-2742.

20) Leslie, A.G., Arnott, S., Chandrasekaran, R., Ratliff, R. L. *J. Mol. Biol.* **1980**, *143(1)*, 49–72.

21) Tan, Z. J. and Chen, S. J. *Biophys J*. **2006**, *90(4),* 1175–1190.

22) Hud, N. V., Sklenar, V., Feigon, J. *J. Mol. Bio*., **1999**, *286 (3),* 651.

23) Dervan, P. B. *Bioorg. Med. Chem.* **2001**, *9*, 2215–2235.

24) Dervan, P. B. *Science*, **1986**, *232(4749)*, 464-471.

25) Pandya, P., Islam, M., Kumar, S., Jayaram, B., Kumar, S. *J. Chem. Sci*., **2010**, *122(2),* 247–257.

26) Drew, H. R., Dickerson, R. E. *J. Mol. Bio.*, **1981**, *151 (3)*, 535-556.

27) Kopka, M. L., Fratini, A. V., Drew, H. R., Dickerson, R. E. *J. Mol. Bio.*, **1983**, *163 (1)*, 129-146.

28) Egli, M., Tereshko, V., Teplova, M., Minasov, G., Joachimiak, A., Sanishvili, R., Weeks, C. M., Miller, R., Maier, M. A., An, H., Cook, P., Manoharan, M. *Biopolymers,* **1998**, *48(4),* 234-52.

29) Lan, T., McLaughlin. L. W. *J. Am. Chem. Soc.*, **2000**, *122*, 6512.

30) Lan, T., McLaughlin, L. W. *Biochemistry*, **2001**, *40*, 968-976.

31) Crick, F. H. C. *J. Mol. Bio.* **1966**, *19 (2)*, 548-555.

32) Ozaki, H. and McLaughlin, L. W. *Nuc. Acids Res.,* **1992**, *20(19),* 5205-5214.

33) Zuckermann, R., Corey, D., Schultz, P. *Nuc. Acids Res.*, **1987**, *15(13),* 5305 – 5321.

34) Mineno, J., Ishino, Y., Ohminami, T., Kato, I.. *DNA Seq.* **1993**, *4(3)*, 135-41.

35) M. Sameiro T. Gonçalves. *Chem. Rev.* **2009,** *109,* 190–212

36) Conway, N. E., Fidanza, J. A., McLaughlin, L. W. *Phosphorus, Sulfur, and Silicon.* **1990**, *51/52*, 27-30.

37) Ganguly, M., Wang, R., Marky, L. A., Gold, B. *J. Am. Chem. Soc.*, **2009**, *131 (34),* 12068–12069.

38) Katritzky, A. R. *Advances in Heterocyclic Chemistry*, Vol 89. ; Academic Press, 2005; p. 240.

Chapter 2: Development of New Protecting/Directing Groups for Glycosylation Reactions in Efforts to Synthesis 3-deaza-3-modified-2′-deoxyadenosine Analogues

2.1 Chemical Synthesis of Nucleic Acids

Deep investigation of DNA structure and stability requires the ability to modify the structure of a nucleic acid. Some modified nucleic acids exist in nature but in order to use these molecules as probes, larger amounts than a cell can produce are necessary. Both non-natural and natural nucleoside analogues can be accessed in sufficient quantities through organic chemistry. The key step in any nucleoside synthesis is the coupling of a heterocycle to a sugar. Often referred to as a glycosylation, the reaction involves coupling a modified heterocycle to a protected sugar. The reaction poses the possibility of 4 products including stereoisomers of the ribose and regioisomers of the heterocycle (**Figure 1**). The exact procedure of this step varies according to the type of heterocycle

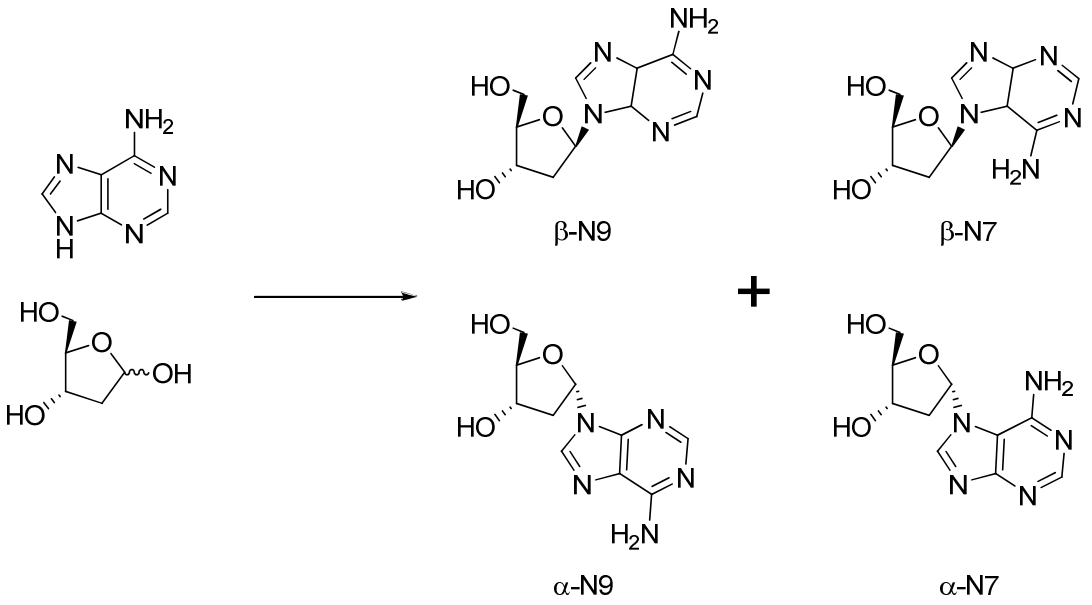
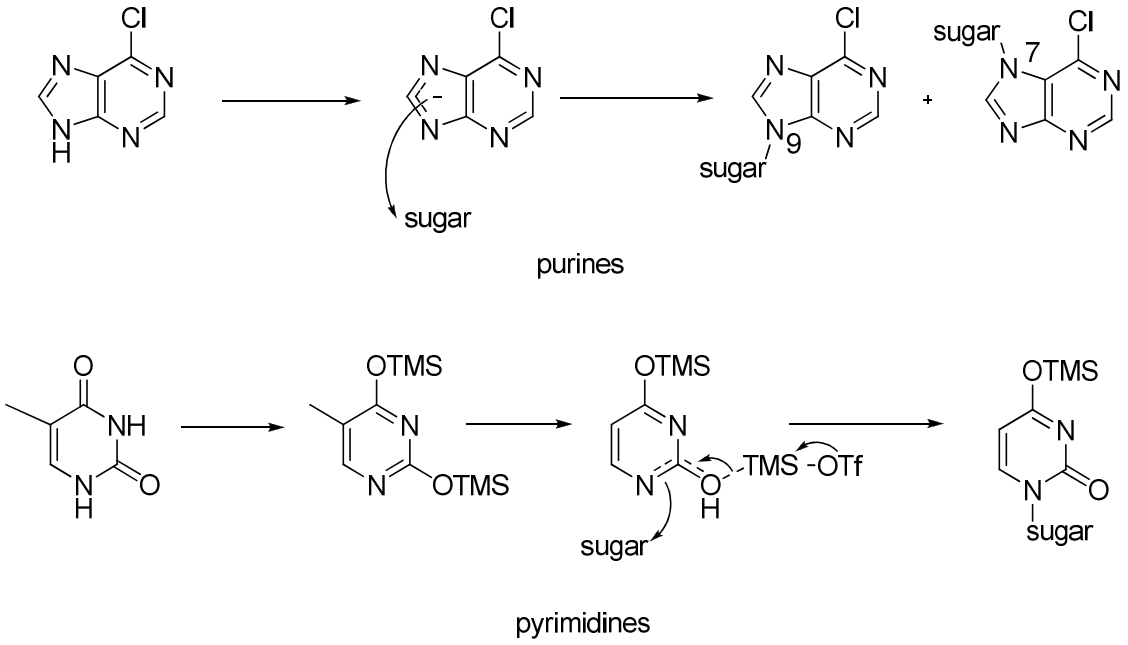


Figure 1. Possible Glycosylation Products as shown with Adenine and 2-deoxyribose.

(purine or pyrimidine) and sugar (ribose or 2-deoxyribose). In general, 6-chloropurines are activated towards glycosylation by deprotonation, while pyrimidines are first silylated and then activated with triflate anions (**Scheme 1**). Naturally occurring nucleosides are β-N9 for purines and β-N1 for pyrimidines. Directing reactions towards these products are important factors in efficiently synthesizing nucleosides and will be discussed throughout this chapter.

Scheme 1. Activation of Purines and Pyrimidines Towards Glycosylation

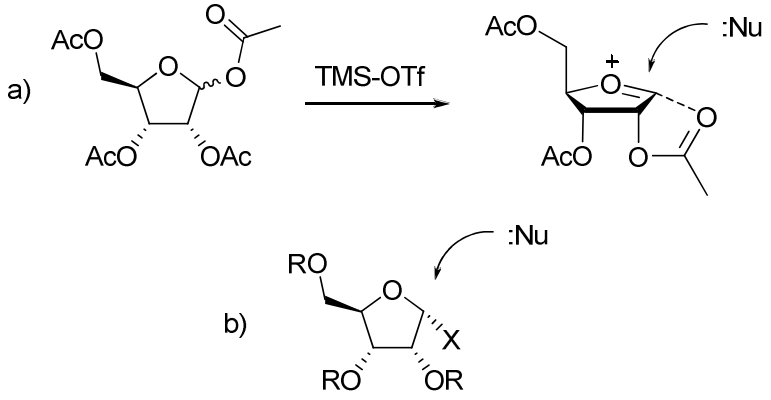


Purine glycosylations result in two regioisomeric products: N9 and N7. N7 purine nucleosides have shown therapeutic potential,[1, 2] but in nature, nucleic acids have N9 conformation, so this is the desired isomer for these studies. Typical ratios of N9 to N7

for purine glycosylations are 5:1,[3] but deviate once substitutions to the heterocycle are considered. Pyrimidines are directed to the proper N1-glycosidic bond since N3 is flanked by 2 electron withdrawing oxygens.

The chemistry of the glycosylation differs depending on whether ribo- or deoxyribo- nucleosides are the target. Standard carbohydrate couplings occur via an S_N1 mechanism. Linkages can be directed to a single diastereomer by a phenomenon referred to as "neighboring group effect". This effect utilizes esters on the 2-hydroxyl to block one face of the planar oxonium ion (**Scheme 2a**).[4] Because 2-deoxyribose lacks this directing hydroxyl functionality, an S_N2 type reaction must take place. Deoxyribose is functionalized with a leaving group in the α- position, so inversion ensures the desired β-diastereomer is formed (**Scheme 2b**).

Scheme 2. Glycosylation using a) protected ribose b) protected, α-deoxyribose



2.2 Synthetic Targets and Previous Routes

As discussed in Section 1.4, 3-deazapurine nucleosides are a class of nucleoside analogues that possess anti-inflammatory and anti-viral properties. These molecules can also function to probe the structure of DNA and how molecules bind to the minor groove. Synthetic access to the minor groove poses a challenge and is partly responsible for the observation that these analogues have not been used more frequently. In order to investigate minor groove binding properties, a series of 3-deaza-3-substituted-2'-deoxyadenosine analogues were designed (**Figure 2**). The first, 3-deaza-2'-deoxyadenosine works to effectively remove hydrogen bonding sites of water molecules within the minor groove. In order to observe how small molecules bind to the minor

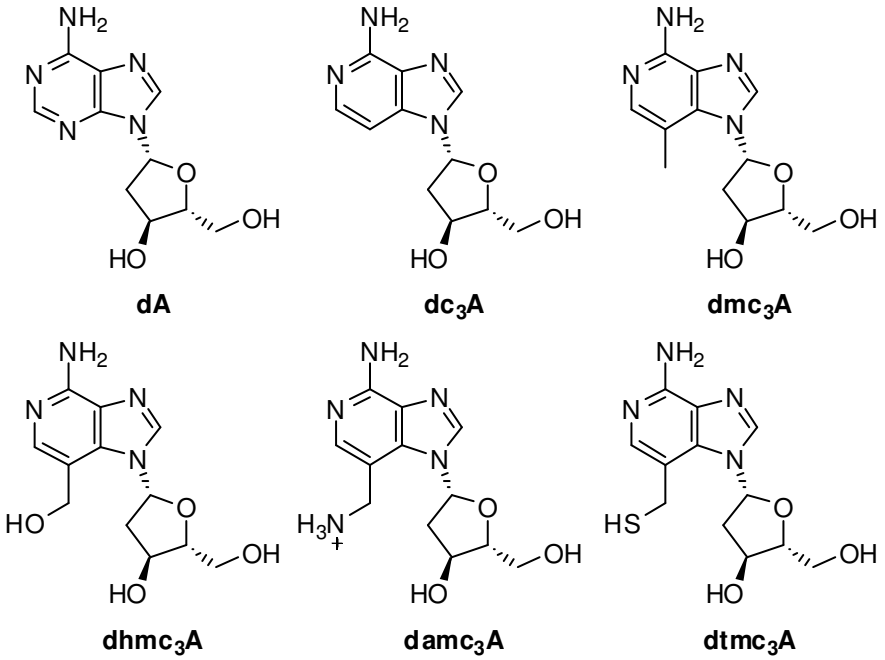
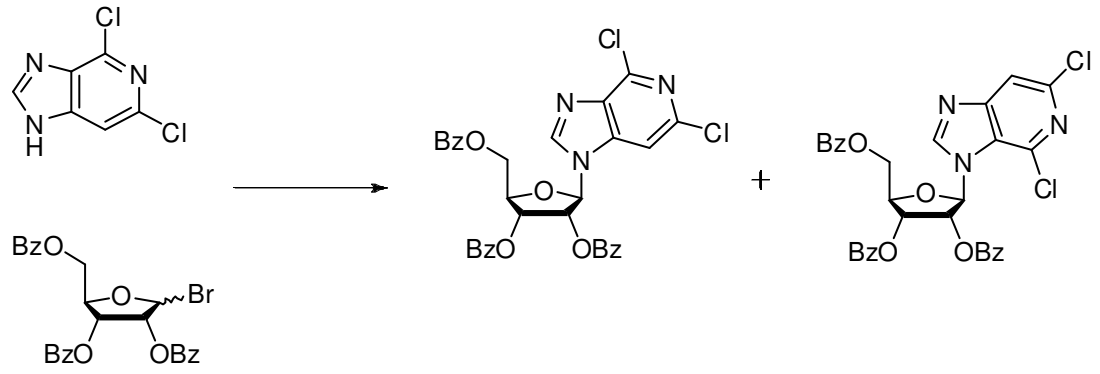


Figure 2. Modifications of dA utilizing dc_3A as a stepping stone for more complex substitutions (dmc_3A, $dhmc_3A$, $damc_3A$, $dtmc^3A$) to study small molecule binding to the minor groove of DNA.

groove, the tethering of hydroxyl group as seen in 3-deaza-3-hydroxymethyl-2'-deoxyadenosine (dhmc$_3$A) and tethering of an amine as shown in 3-deaza-3-aminomethyl-2′-deoxyadenosine (damc$_3$A) were designed (Chapter 3). Lastly, in conjunction with damc$_3$A, 3-deaza-3-thiomethyl-2′-deoxyadenosine (dtc$_3$A) was designed to study site specific labeling with fluorophores in the minor groove (Chapter 4). To efficiently quantitate the effects of tethered functional groups, 3-deaza-3-methyl-2′-deoxyadenosine (dmc$_3$A) was designed to approximate the steric bulk these groups introduce into the minor groove of DNA.

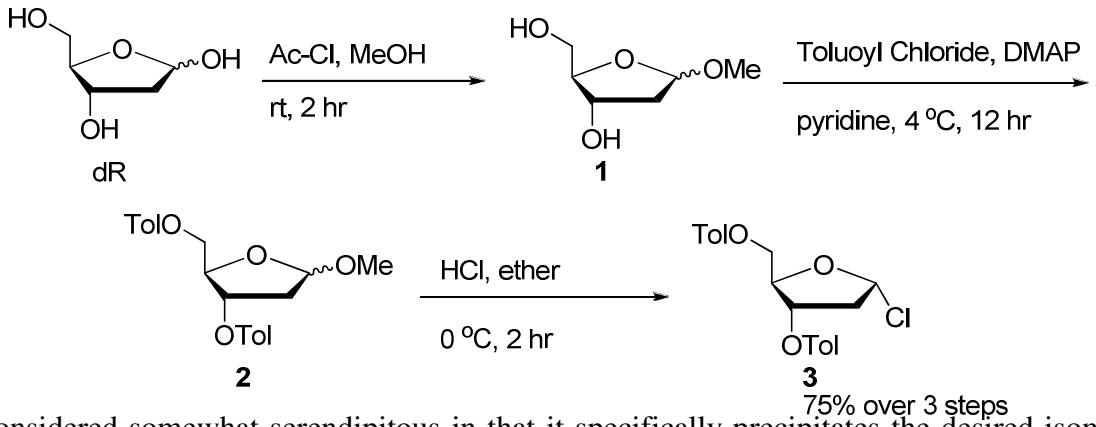
The first example of synthetic 3-deazaadenosine nucleosides in the literature comes from May and Townsend in 1973.[5, 6] Convergent synthesis of the analogue occurred between the coupling of the heterocycle (4,6-dichloro[4.5-*c*]pyridine) and a protected ribose sugar (2,3,5-tri-O-benzoyl-D-ribofuranosyl bromide) (**Scheme 3**). It was 2 decades later that the deoxyribose form of 3-deazapurine nucleosides was synthesized by Matsuda and Minakawa.[7] This synthesis began with 4-carboxamidoimidazole nucleosides with formylmethyl or cyanomethyl groups in position 5. Problems with these syntheses relate to anomeric and regioisomeric mixtures.

Scheme 3. First Convergent Synthesis of Protected dc$_3$A.



Most procedures for purine glycosylations to 2-deoxyribose follow an S_N2 coupling of an anionic heterocycle with 3,5-protected, α-chloro sugar. Heterocycles can be synthesized to possess desired functionality before glycosylation and will be discussed in detail. The synthesis of the sugar begins with 2-D-deoxyribose (**Scheme 4**).[8] The anomeric center is first protected as a methyl ether (**1**), followed by toluoyl protection of the 3- and 5-hydroxyls (**2**). The last step involves displacement of the methyl ether with hydrochloric acid to afford the α-chloro isomer as a precipitate. The precipitation step is

Scheme 4. Synthesis of 1-a-2-deoxy-3,5-O-*p*-toluoyl-D-ribosylchloride (**3**).

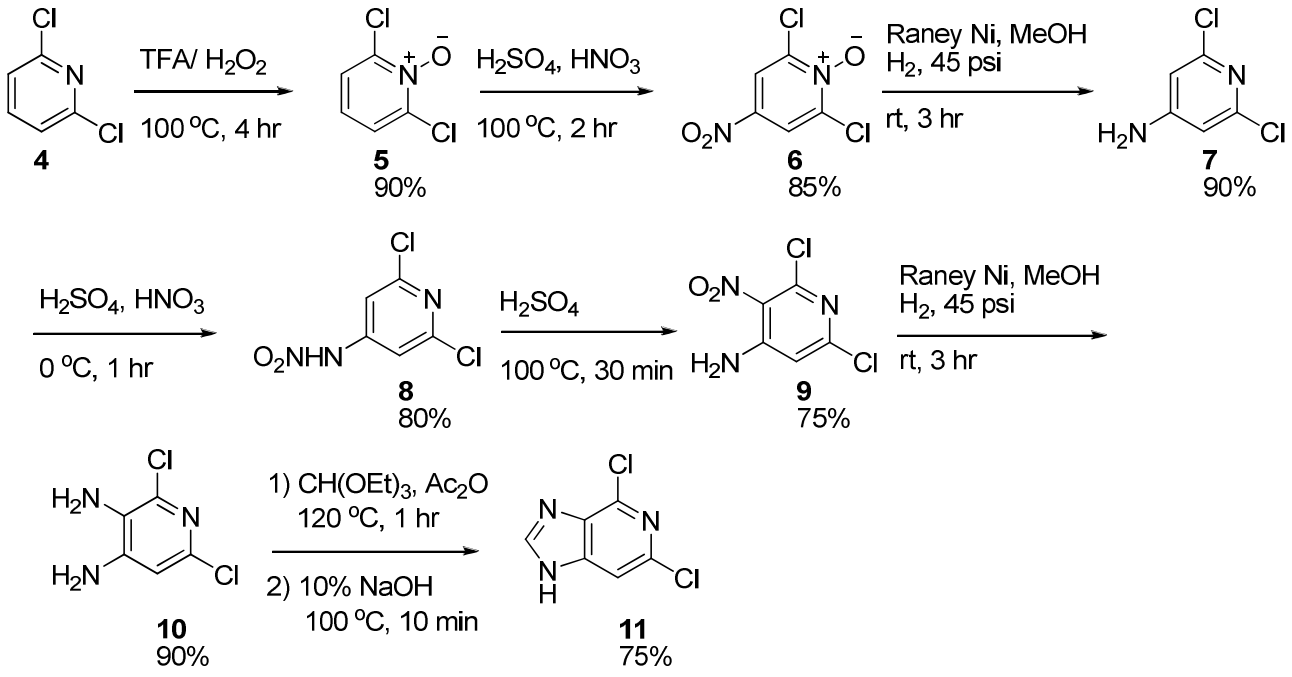


considered somewhat serendipitous in that it specifically precipitates the desired isomer and the undesired isomer remains soluble. This reaction funnels all intermediates to the α-chloro anomer since the β-anomer is in equilibrium with starting materials. Other protecting groups for the 3- and 5-hydroxyls were used, such as acetyl or benzoyl, but only toluoyl protecting groups afforded the desired isomer via precipitation.

The most widely accepted synthetic route for 3-deaza-2'-deoxyadenosine utilizes this sugar and a 6-chloro-3-deazapurine.[9-12] Mizuno used 6-chloroimidazo[4.5-c]pyridine

as the heterocyclic synthon, but 2,6-dichloroimidazo[4.5-*c*]pyridine was later adopted as the norm by Cosstick and coworkers. The electron density of deazapurine systems is drastically different than the parent purine system due to replacement of this nitrogen with C-H. Particularly, 6-chloro-3-deazapurine nucleosides have a high electron density on C6, making nucleophilic displacement with ammonia to afford 3-deazaadenosine extremely slow and low yielding. Addition of an electron-withdrawing chlorine in the 2-position decreases the electron density on C-6 and allows the displacement to take place, albeit with less than desirable yields. The 2-chloro substituent can be removed from the heterocycle by reduction to generate the target 3-deaza-2'-deoxyadenosine. Initially, we followed precedent to synthesize 2,6-dichloro-3-deazapurine.[13] The use of a novel protecting/directing group for glycosylations required use of 6-chloro-3-deazapurine; that synthesis will be discussed in section 2.3. The dichloro-heterocycle is synthesized from 2,6-dichloropyridine in 28% yield over 8 steps (**Scheme 5**). Oxidation of the starting

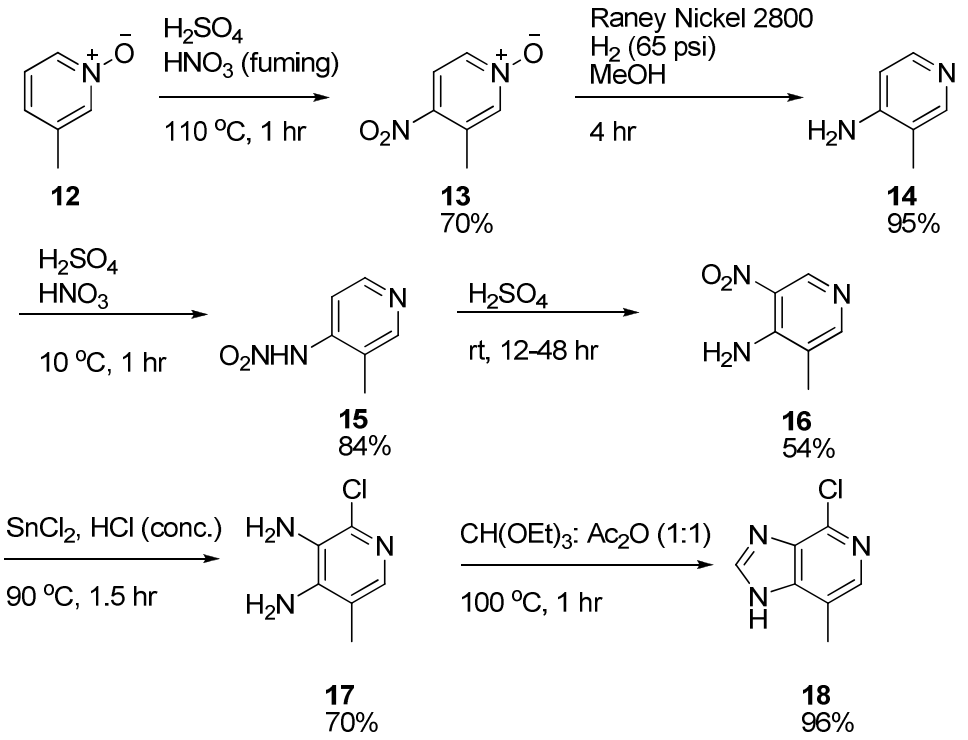
Scheme 5. Synthesis of 2,6-dichloro-3-deazpurine



material **4** followed by nitrating conditions resulted in installation of the future N9 position of the heterocycle **6**, which was then reduced with Raney Ni and H_2 in a Parr shaker to afford **7**. Since strong nitrating conditions led only to decomposition, the second nitration was performed at 0 ^{o}C and then followed with an acid-catalyzed rearrangement to install the N7 position (**9**). To complete the 3-deazapurine ring system, the nitro group is reduced again with Raney Ni and then reacted with triethylorthoformate in the presence of acetic anhydride to cyclize the imidazole moiety. Reflux with diluted sodium hydroxide followed by neutralization with hydrochloric led to precipitation of 2,6-dichloro-3-deazapurine (**11**).

Synthesis of the methyl-substituted deazapurine begins with 3-picoline-*N*-oxide (**Scheme 6**, **12**).[14] Nitration with concentrated H_2SO_4 and fuming HNO_3 gave **13** which was followed by reduction of the nitro group and concomitant removal of the *N*-oxide with Raney Ni/H_2 in a Parr shaker at 65 psi to give **14**. Similar to steps in the synthesis of 2,6-dichloro-3-deazapurine, a second nitration with external cooling was performed to give nitroamine **15**, which was followed by sulfuric acid catalyzed rearrangement. Contrary to the previous synthesis, this reaction was not reproducible, sometimes complete in 8 h and sometimes requiring 48 h or more. The following reduction step was also problematic. The reported yield of the reduction-chlorination of **16** to **17** was 29% with numerous side products. The optimized conditions used more $SnCl_2$ (4.9 vs. 1.6 equiv), 5 mL/g of HCl, and reproducibly gave 70% isolated yield without the need for chromatography. Finally, cyclization to **18** using triethylorthoformate and acetic

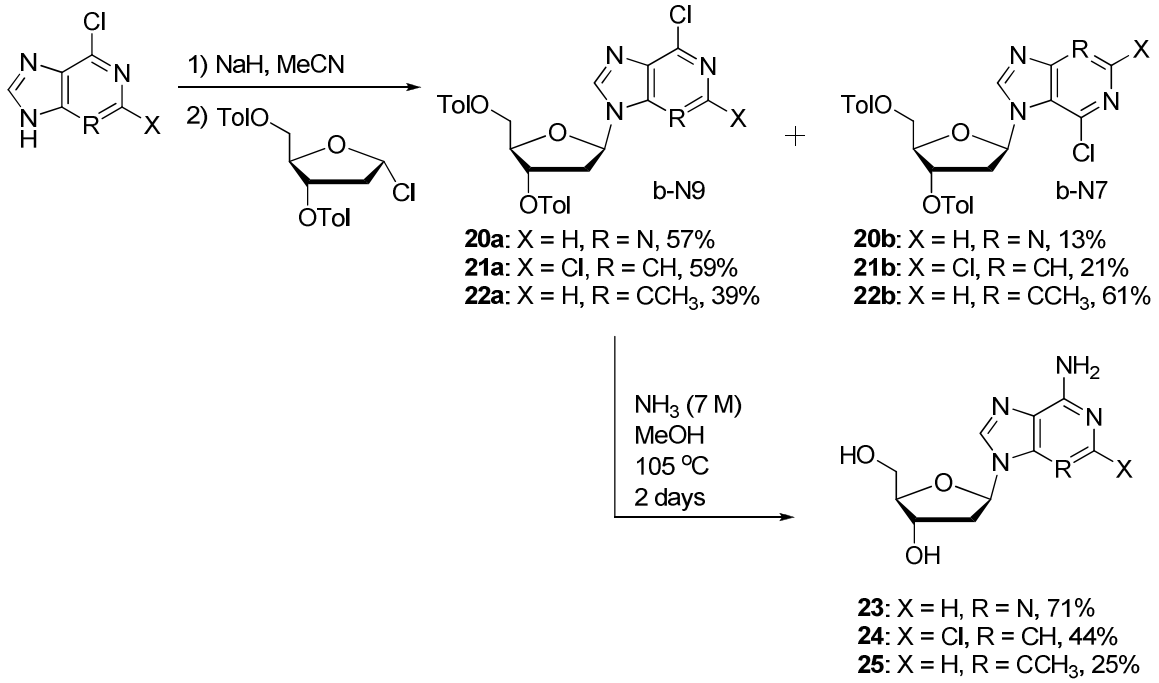
Scheme 6. Synthesis of 6-chloro-3-deaza-3-methylpurine.



anhydride afforded **23** in 96% with purification possible by passing the crude product through a silica plug.

Protected nucleosides of dA, dc_3A, and dmc_3A were produced by glycosylation vivia sodium hydride deprotonation followed by addition of α-chlorosugar **3** (**Scheme 7**) of 6-chloropurine (19) and the 3-deazapurine derivatives (**11** and **18**). The coupling of 6-chloropurine **19** and 2,6-dichloro-3-deazapurine **11** resulted in similar yields of the desired β-N9 nucleoside (57% and 59%, respectively), but addition of the methyl group on the 3-position of 6-chloro-3-deaza-3-methylpurine **18** drastically shifted the product to β-N7 (61%). The protected nucleosides (**20a, 21a, 21b**) are reacted in an enclosed vessel with concentrated methanolic ammonia to complete the synthesis of the free nucleosides in yields of 71% for dA, 44% for dc_3A, and 25% for dmc_3A. The overall yield for the

Scheme 7. Formation of dA, dc_3A, and dmc_3A.

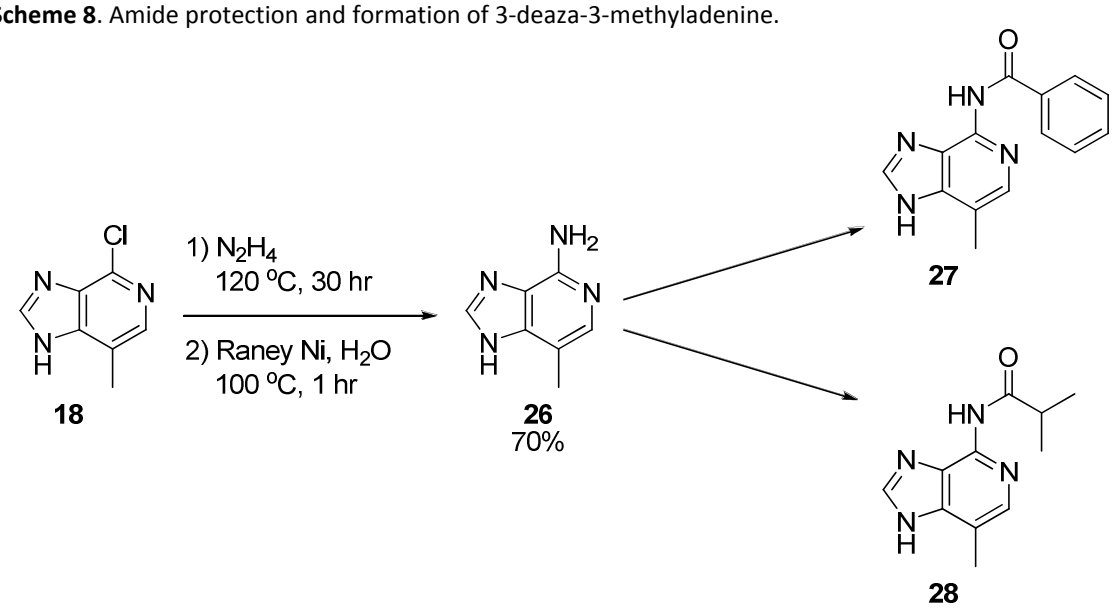


synthesis of dA (**23**) from 6-chloropurine **19** is 41%. It is clear that the 3-deazapurine system is less reactive towards nucleophilic displacement as the yield of **24** from **11** is 26% over two steps. Lastly, the overall yield of the methyl derivative, dmc_3A, over two steps is 10%, which stems from both introduction of steric bulk into the 3-position during glycosylation and poor reactivity of the heterocycle to nucleophilic displacement. The methyl group effectively interferes with glycosylation at the N9 of the heterocycle, driving the product towards a β-N7 nucleoside. It was clear from these results that the introduction of larger substituents in the 3-position necessary for the syntheses of hydroxymethyl, aminomethyl, and thiomethyl derivatives would only further decrease the yield of β-N9 product from the glycosylation.

2.3 Directing and Protecting Groups for Glycosylation Reactions

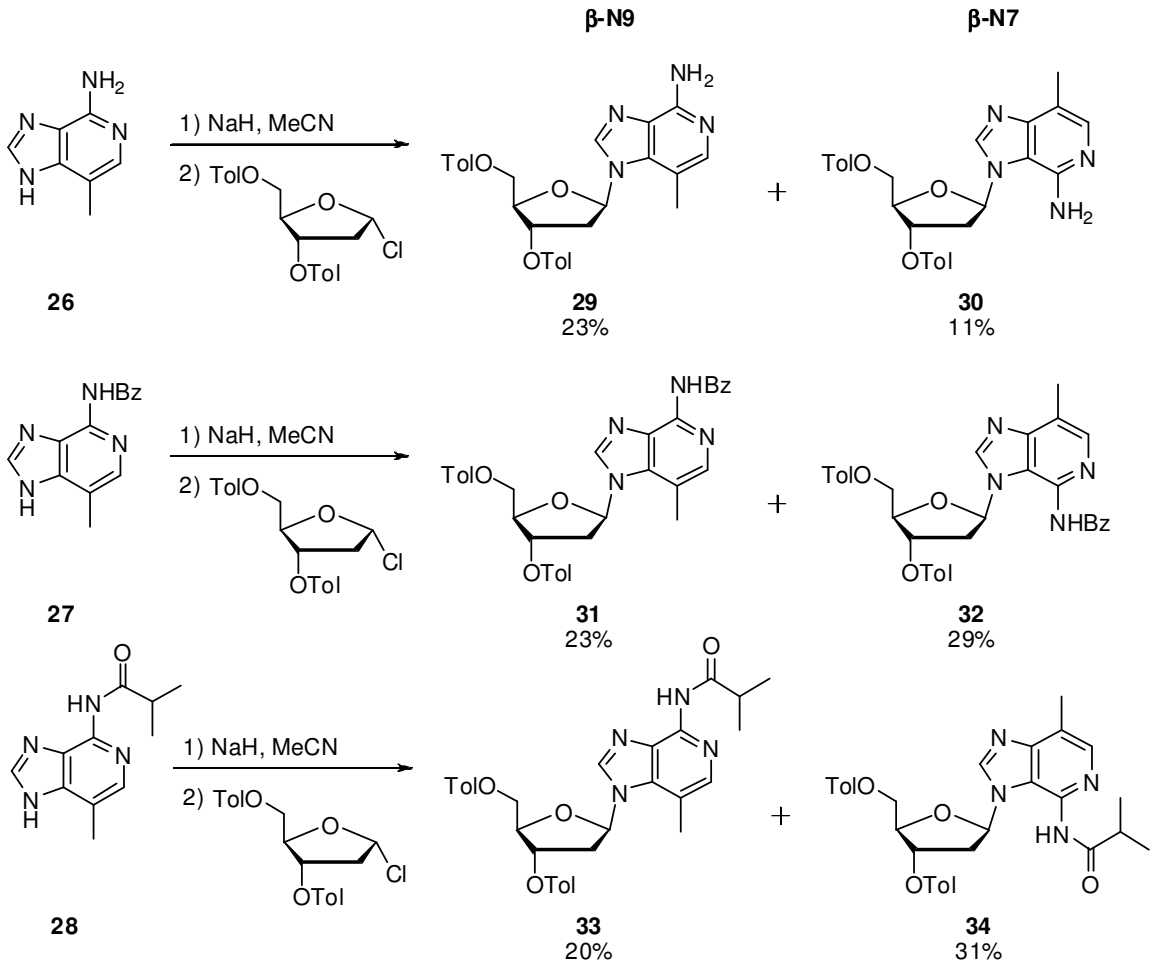
Glycosylation reactions have commonly been referred to as the "bottle-neck" step[15] of any nucleoside synthesis due to their less than desirable yields and multiple products. In addition, 3-deazapurines require harsh conditions to install the N^6-exocyclic amine,[16] which is reflected in the yield of this step. Our first goal, in collaboration with Joseph W. Arico and Amy K. Calhoun, was to install the amine on the heterocycle prior to glycosylation, as heterocycles are more stable than nucleosides to robust conditions. Our second goal involved design of a protecting group that functions to both protect the amine and direct the glycosylation to β-N9 since leaving the exocyclic amine unprotected during glycosylation leads to unwanted side reactions with NaH. Our first attempt used amide groups to protect the exocyclic amine on 3-deaza-3-methyladenine (**Scheme 8**). Refluxing 6-chloro-3-deaza-3-methylpurine **18** with anhydrous hydrazine, followed by Raney Nickel reduction afforded 3-deaza-3-methyladenine **26**, which could then be

Scheme 8. Amide protection and formation of 3-deaza-3-methyladenine.



protected. Benzoyl and isobutyroyl protecting groups are commonly used on the exocyclic amines of nucleoside phosphoramidites as they are easily cleaved in concentrated ammonium hydroxide at 55 ^{o}C. Reaction of **26** with benzoyl chloride or isobutyrl anhydride resulted in amides **27** and **28**, respectively. These and unprotected amine **26** were subjected to standard glycosylation conditions (**Scheme 9**), but each derivative gave lower yields that the initial reaction with the chloro substituted analogue.

Scheme 9. Glycosylations utilizing amine and amide functionalities.

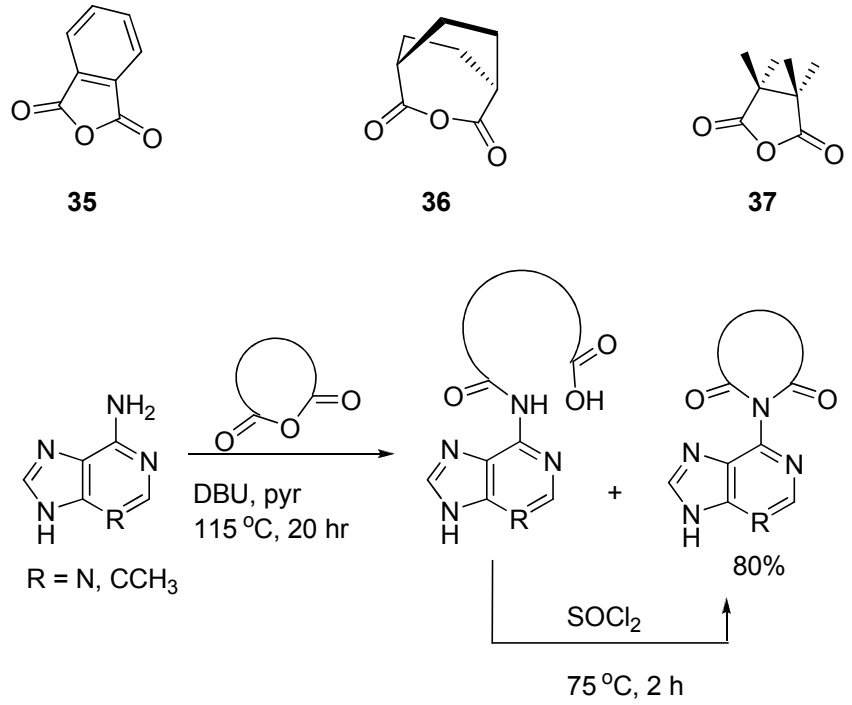


About this time, the Robins group published a series of manuscripts detailing the use of substituted imidazole and triazole protecting groups that worked to direct the glycosylation ratio.[17-19] The azoles were installed by treatment of 6-chloropurine in DMF and could be removed after the glycosylation to expose the N^6-amine. The total yield of products from use of 2-propylimidazole as the directing group was 50%, with a 4:1 ratio of β-N9 to β-N7 coupling products. The yield of β-N9 products had not improved, but the 10% yield of β-N7 has significantly decreased, which should make separation of the two isomers easier. Regrettably, no reaction took place upon treatment of the deazapurines with 2-propylimidazole, due to the electron-rich properties of 3-deazapurines.

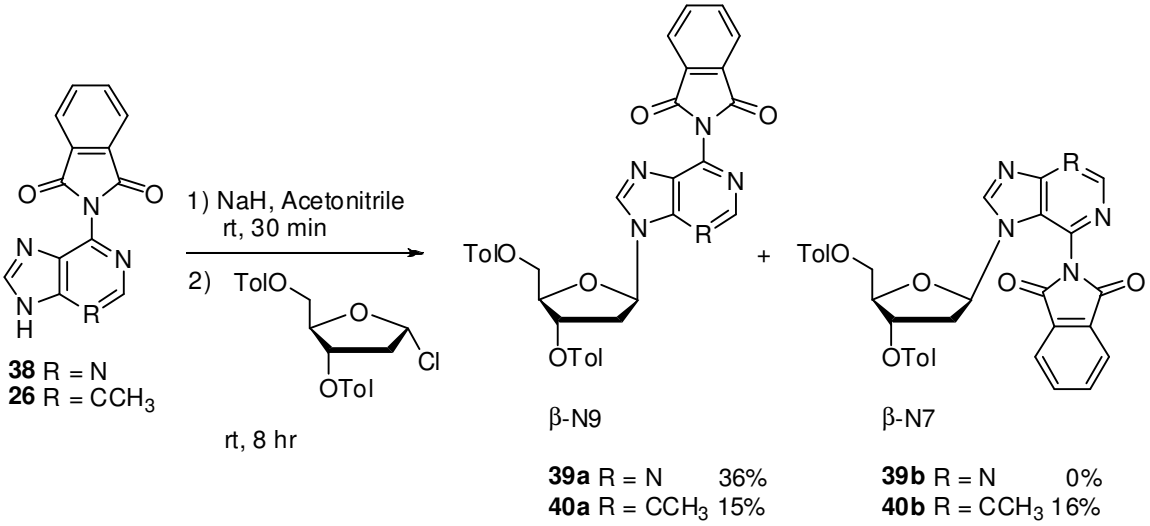
The final endeavor involved the use of cyclic imides to protect the amine and direct the glycosylation. In general, the cyclic imides could be installed by mixing the corresponding anhydride with an aminopurine derivative in the presence of DBU and pyridine at reflux for 20 hours (**Scheme 10**). Trace amounts of amide co-eluted with the imide during column chromatography and could be converted by treating the mixture with thionyl chloride. The fully cyclized imide could be obtained in approximately 80% yield.

We began our studies with phthalic anhydride **35** as it is commercially available and has been used as a protecting group in phosphoramidite chemistry. Glycosylation of phthalimide protected 6-aminopurine **38** resulted in 36% β-N9 as the sole product, suggesting that this group effectively blocks the N7 position from reacting (**Scheme 11**).

Scheme 10. Imide Protection of 6-aminopurines



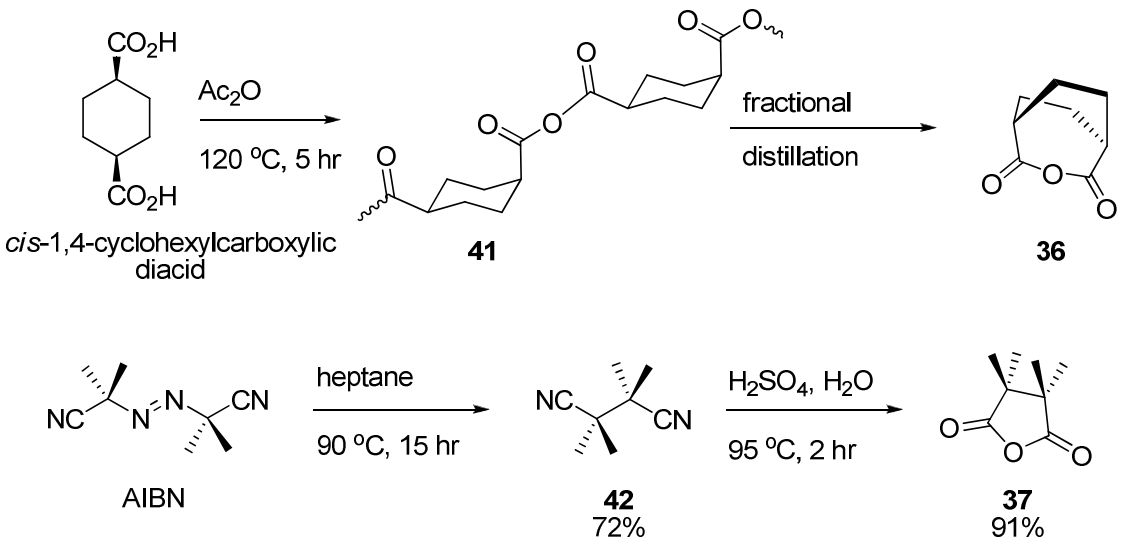
Scheme 11. Glycosylation of Phthalimide Protected Purines



Using the 3-deaza-3-methyl protected analogue, the total yield of glycosylation was 60% with the ratio of β-N9 to β-N7 as 1:1, indicating that phthalimide cannot overcome the sterics presented by the methyl group. In addition, an equal amount of α- and β- isomers were produced, leaving only 15% total yield of the desired product. It should be noted that the protected sugar should only yield β-anomer coupling products, but formation of α-anomer products still occurs. Glycosylation reactions require polar, aprotic solvents. Sugars are capable of anomerizing in polar solvents after some time, especially with the long coupling times used for this reaction. Shorter coupling times have shown to reduce the formation of α-products without much affect on the yield of β-N9 product, but currently the heterocycles are only soluble in polar solvents so anomerization remains problematic. Further study of this protecting group revealed a particular lack of stability – it could be effectively removed in 3 M NH_3 in methanol in under 20 minutes. For some applications, such as post-DNA synthesis deprotection, this characteristic may be advantageous, but for our purposes it was not stable enough to use as a protecting group to synthesize phosphoramidites.

We next focused our efforts on the other two potential groups. As aliphatic imides, it seemed promising that these tetrahedral protecting groups would serve to block the N7-position more effectively than the planar phthalimide group. Anhydrides **36** and **37** are not commercially available and therefore must be produced from corresponding available materials. *Cis*-1,4-cyclohexyldicarboxylic acid could be used to obtain **36** after treatment with acetic anhydride followed by distillation (**Scheme 12**).[20]

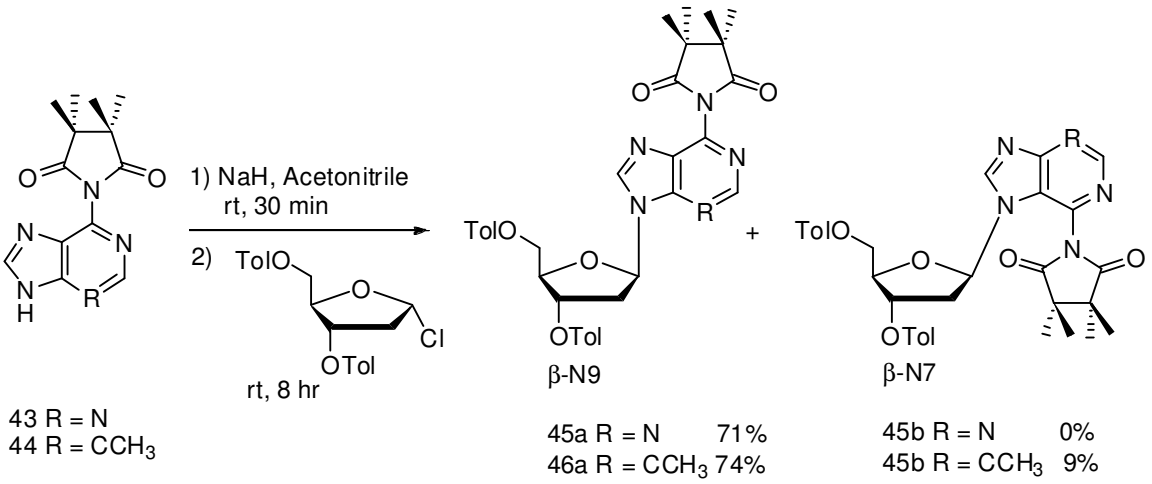
Scheme 12. Preparation of Aliphatic Cyclic Anhydrides



Tetramethylsuccinyl anhydride **37** was prepared from azobisisobutyronitrile (AIBN) as its commercial source. AIBN is typically used in small amounts as an initiator in radical chemistry; we were able to take advantage of its ability to fractionate and terminate with itself to form intermediate **42** in 72% yield. Treatment of **42** with diluted sulfuric acid hydrolyzed the dinitrile to the corresponding di-acid which could then be dehydrated to form cyclized anhydride **37**.[21] With both aliphatic anhydrides on hand, we attempted to protect the exocyclic amine on 6-aminopurine and 6-amine-3-deaza-3-methylpurine.

Cis-1,4-cyclohexyl anhydride **36** proved fruitless in attempts to install the imide on the heterocycle, most likely due to its tendency to polymerize at high temperatures, as seen in its synthesis. Tetramethylsuccinyl anhydride **37** followed the general protocol (**Scheme 10**) to protect the exocyclic amine. Glycosylation of the protected nucleobases

Scheme 13. Glycosylation of M4SI-protected Purines



resulted in the following: tetramethylsuccinyl (M_4SI-) adenine resulted in exclusive β-N9 product in 71% yield, while the protected methyl derivative resulted in 74% β-N9 with 9% β-N7 (**Scheme 13**). The results of the glycosylation reactions with chloro-, phthalimide, and M4SI- protected heterocycles are highlighted in **Table 1**.

In order to investigate the orientation of M_4SI- in respect to the nucleobase, a crystal structure of **43** was obtained **(Figure 3**). Our initial instincts with the cyclic imide led us to believe the carbonyls of the protecting group would remain in the plane of the nucleobase as to maximize p-orbital electron sharing with the aromatic purine. To our surprise, the protecting group was rotated 52^{o} relative to the plane of the heterocycle, most likely due to the proximity of the imide oxygen to C5 and N7 of the heterocycle (3.0 and 3.1 Å, respectively). The methyl groups can be observed positioned "umbrella-like" over the N7 position, likely the culprit of this group's ability to direct the glycosylation to

β-N9. This also helps explain why phthalimide worked so poorly as a directing group with the methyl derivate (**Table 1**). It appears that larger substitutions on the succinimide group might function to completely block formation of β-N7 nucleosides, leading to a new class of protecting and directing groups for purine glycosylation.

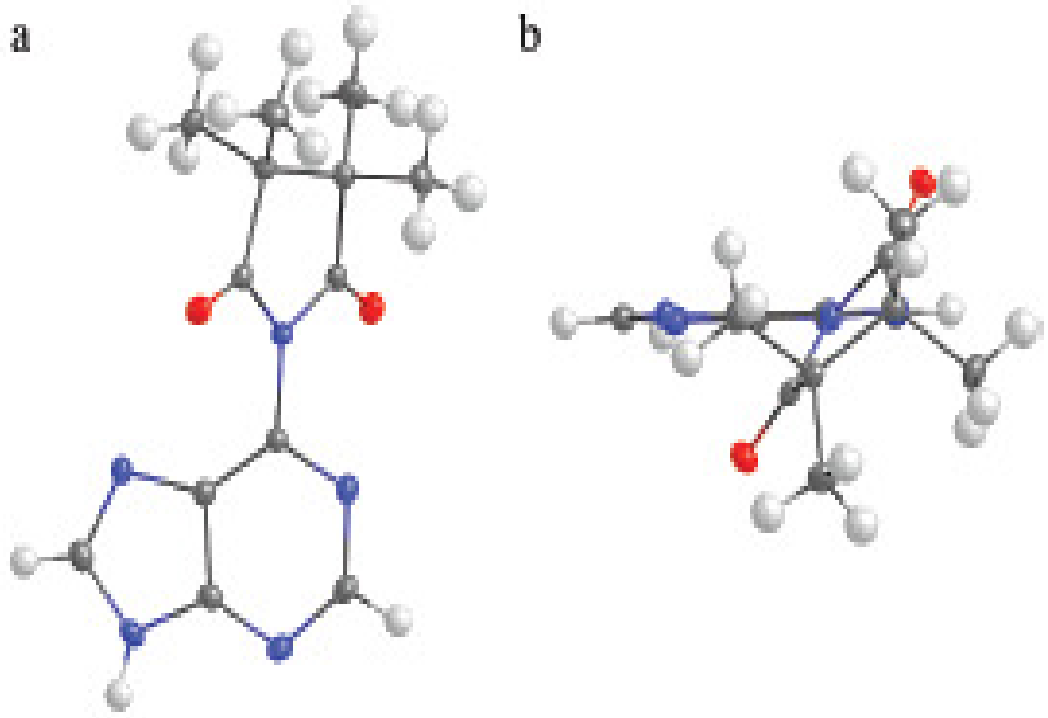
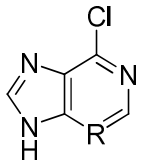
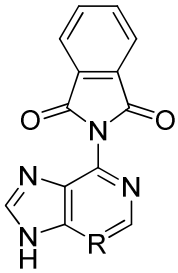
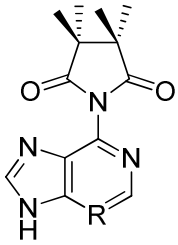


Figure 3. Crystal Structure of M_4SI-protected adenine. a) Face-On View. b) Top-Down View showing methyl positioned over N7.

Table 1. Glycosylation Results of Various 6-substituted purines

Heterocycle	R	β-N9	β-N7	β-N9:β-N7	α-N9/α-N7
	**19** R = N	57%	13%	4.7 : 1	0%
	18 R = CCH_3	39%	69%	1 : 2.2	0%
	**38** R = N	36%	0	>99 : 1	0
	26 R = CCH_3	15%	16%	1 : 1.1	31%
	**43** R = N	71%	0%	>99 : 1	6%
	44 R = CCH_3	74%	9%	8.2 : 1	3%

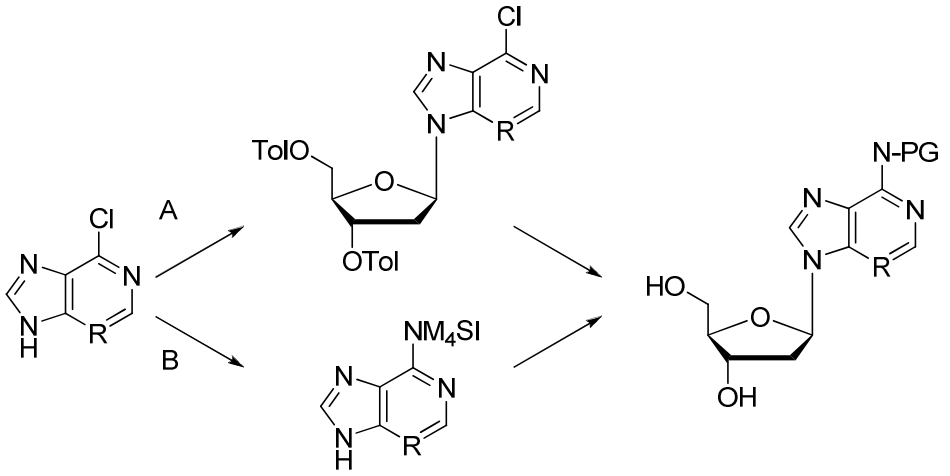
2.4 Next Generation Protecting and Directing Groups

As described in the previous section, purine glycosylations have served as a laborious step in purine nucleoside synthesis, making it expensive and time consuming. Our work with tetramethylsuccinimide (M_4SI-) has improved the overall yield of N^6-protected adenosine from 6-chloropurine derivates by as much as 23% (**Table 2**). However, problems still exist with the current synthesis, particularly to syntheses involving substitutions at the 3-position of the purine. Separation of unwanted isomers produced during the glycosylation requires very long and tedious chromatography techniques, including extra long columns, special silicia gel, and multiple columns in order to completely isolate β-N9 from β-N7, α-N9, and α-N7 isomers.

Next generation protecting and directing groups are targeted towards resolving these issues by complete blockage of the N7 position and less polar reaction solvents. From our results with tetramethylsuccinimide as a protecting and directing group, we designed an analogue containing two cyclohexyl groups in place of the four methyls on the anhydride in collaboration with Ayan Pal. These cyclohexyl groups serve two functions: 1) to increase the steric size of the substitutions on the protecting group as to direct glycosylation solely to N9 product and 2) to increase the aliphatic character of the heterocycle as to make it more soluble in less polar solvents, such as dichloromethane, which should help to prevent α-anomeric products from forming. The synthesis of the dicyclohexylsuccinyl anhydride and protection of 6-aminopurine are analogous to that the

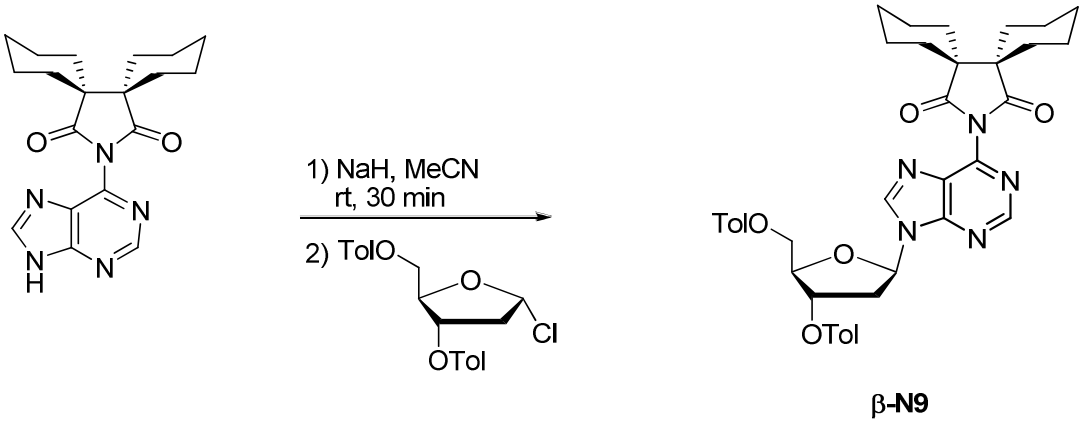
tetramethylsuccinyl anhydride. Use of this derivative under standard glycosylation shows 90% yield of β-N9 nucleoside as the sole product.

Table 2. Comparisons of Total Yield of N^6-Protected Adenosine Derivatives from corresponding 6-chloropurines



R	Pathway A	Pathway B
N	32%	40%
CH	17%	31%
CCH_3	8%	31%

Scheme 14. Next Generation Protecting Groups.



β-N9

2.5 Experimental

General Procedures

All reactions were carried out under an inert atmosphere with dry solvents under anhydrous conditions unless noted otherwise. Dry tetrahydrofuran (THF), diethyl ether (Et_2O), *N*,*N*-dimethylformamide (DMF), pyridine (pyr), acetonitrile (MeCN), and dichloromethane (DCM) were obtained by passing commercially available pre-dried, oxygen-free formulations through activated alumina columns. Dry methanol (MeOH) was obtained by distillation from $Mg(OMe)_2$. Yields refer to chromatographically and spectroscopically (^{1}H NMR) homogeneous materials, unless otherwise stated. Reagents were purchased at the highest commercial quality and used without further purification, unless otherwise stated. Reactions were monitored by thin-layer chromatography (TLC) carried out on 0.25 mm Silicycle TLGR10011B-323 60 Å plates using UV light as visualizing agent and sulfuric acid stain and heat as developing agents. Dynamic Adsorbents silica gel (60 Å, particle size 32-63 μm) was used for flash column chromatography. Difficult separations were carried out using "chromatospec silica gel" (E. Merck Chromatospec silica gel (60 Å, particle size 15-40 μm). NMR spectra were recorded on Varian VNMRS 400, VNMRS 500, VNMRS 600, or INOVA 500 instruments and calibrated using residual undeuterated solvent ($CDCl_3$: δ H = 7.26 ppm, δ C = 77.16 ppm, acetone-*d*6: δ H = 2.05 ppm, δ C = 29.84, DMSO-*d*6: δ H = 2.50 ppm, δ C = 39.52 ppm, methanol-*d*4: δ H = 3.31 ppm, δ C = 49.00 ppm) as an internal reference. The following abbreviations are used to designate the multiplicities: s = singlet, d = doublet, t = triplet, q = quartet, quin = quintet, m = multiplet, br = broad. High-resolution

mass spectra (HRMS) were recorded on a Waters LCT Classic or JEOL AccuTOF mass spectrometer using ESI (electrospray ionization) or DART (direct analysis in real time).

1-α-2-deoxy-3,5-O-*p*-toluoyl-D-ribosyl chloride (3).[4] Acetyl chloride (210 uL, 2.95 mmol) in MeOH (10 mL) was stirred for 1.5 hr at room temperature under N_2 atmosphere. To the solution was added 2-deoxy-D-ribose (10 g, 74.6 mmol). The solution was allowed to stir for 2 hr at room temperature until TLC indicated complete consumption of starting material and conversion to **1**. $AgCO_3$ was added until the solution was neutral and the suspension was filtered through a celite pad. The filtrate was evaporated to dryness to give a honey-like oil. To this oil was added DMAP (1.13 g, 9.69 mmol) and pyridine (80 mL), followed by addition of *p*-toluoylchloride (20.7 mL, 156 mmol). After 12 hr at 0 ^oC, intermediate **2** was diluted with 150 mL of water and extracted with diethyl ether (3 x 100 mL). The organic layer was then washed with $NaHCO_3$ (5 x 50 mL) and brine (50 mL). The organic layer was collected, dried over Na_2SO_4, filtered, and evaporated to dryness. The final step involves dissolving **2** in diethyl ether (20 mL) and cooling to 0 ^oC. HCl gas is made *in situ* with NaCl and H_2SO_4 in a separate apparatus and then bubbling into the protected sugar solution. After approximately 10 min, a precipitate should form. This precipitate is the α-chloro anomer. If this is not the case, $HCl_{(g)}$ is bubbled through for an addition 20 minutes. If there is still no precipitate, the mixture can be concentrated and then precipitated from concentrated solution in CH_2Cl_2 by slow addition into rapidly stirring Et_2O in an ice bath. ^1H NMR (500 MHz, $CDCl_3$) δ 7.99-7.86 (m, 4H), 7.26-7.20 (m, 4H), 6.45 (d, J = 5.1 Hz, 1H), 5.54

(m, 1H), 4.84 (m, 1H), 4.66 (dd, J = 12.1, 9.0 Hz, 1H), 4.57 (dd, J = 12.1, 7.8 Hz, 1H), 2.85 (m, 1H), 2.73 (d, J = 15.0 Hz, 1H), 2.40 (s, 3H), 2.39 (s, 3H).

4-Nitro-3-picoline *N*-oxide (13)[14] To a flask containing 3-picoline *N*-oxide **12** (20 g, 183 mmol) was slowly added 100 mL of concentrated H_2SO_4. While keeping the temperature below 80 °C, 40 mL of 90% fuming HNO_3 was added. The reaction mixture was then carefully heated to 100 °C until an exothermic reaction was observed with evolution of a brown gas. The flask was removed from heat until the bubbling had subsided, then heated at 120 °C for 2 h, after which it was cooled to room temperature and poured over 300 g of crushed ice. This mixture was placed in an ice bath and titrated with concentrated NH_4OH until a yellow precipitate formed at basic pH. The mixture was filtered and the solid collected and dried to yield 19.74 g (70%) of **13**. This compound could be used directly for the next step; recrystallization from methanol yielded analytically pure material. ^{1}H NMR (400 MHz, acetone-d_6): δ 8.29 (s, 1H), 8.19 (m, 1H), 8.11 (d, J = 12 Hz, 1H), 2.59 (s, 3H); MS (ESI): m/z 154 (M^+).

4-Amino-3-methylpyridine (14)[14] Raney nickel 2800 (8.0 g; **CAUTION**: pyrophoric; weighed under tared volume of water) was added to 100 mL of methanol and pre-reduced in a Parr shaker with H_2 at 60 psi for 5 min. Compound **13** (8.0 g, 52 mmol) was added and reduced at 65 psi for 4 h. The mixture was then filtered through Celite, washed thoroughly with methanol and dried *in vacuo* to obtain 5.34 g (95%) of a slightly green solid. Compound **14** is oxidized easily; therefore it was purged with argon and stored at 4

°C until used. ^{1}H NMR (400 MHz, acetone-d_6): δ 7.96 (s, 1H), 7.94 (d, J = 4.8 Hz, 1H), 6.55 (d, J = 4.8 Hz, 1H), 5.24 (bs, 2H), 2.08 (s, 3H); MS (ESI): m/z 108 (M^+).

4-Nitramino-3-methylpyridine (15)[14] Compound **14** (6.80 g, 63 mmol) was dissolved in 51 mL of concentrated H_2SO_4 and cooled to 5 °C in a dry ice/ethanol bath. 70% HNO_3 (21 mL) was added slowly to maintain the reaction temperature below 10 °C. The reaction mixture was allowed to stir at 10 °C for 1 h and then poured onto 200 g of crushed ice. Concentrated NH_4OH was then added until a creamy off-white precipitate was formed at pH 7. The mixture was filtered and the precipitate recrystallized from hot water to yield 8.11 g (84%) of **15**. ^{1}H NMR (400 MHz, DMSO-d_6): δ 8.15 (d, 2H), 8.00 (d, J = 6.6 Hz, 1H), 2.08 (s, 1H); MS (ESI): m/z = 153 (M^+).

3-methyl-5-nitropyridin-4-amine (16)[14] Compound **15** (2.00 g, 13 mmol) was added to 19 mL of concentrated H_2SO_4 in small portions to avoid decomposition of the starting material. The mixture was allowed to stir at room temperature for 18-48 h until complete (reaction was monitored by removing aliquot, neutralizing with conc. NH_3, filtering, and analyzing the precipitate by ^{1}H NMR) and then quenched by pouring over 200 g of crushed ice. This mixture was then slowly titrated with concentrated conc. NH_3 while cooling in an ethanol/dry ice bath. When the mixture reached pH 6.5–7 a yellow precipitate formed. This was filtered and dried to yield 1.02 g (54%) of **16**. ^{1}H NMR (400 MHz, DMSO-d_6): δ 8.87 (s, 1H), 8.09 (s, 1H), 7.66 (bs, 2H), 2.12 (s, 3H). MS (ESI): m/z 153 (M^+).

2-chloro-5-methylpyridine-3,4-diamine (17)[14] Anhydrous stannous chloride (12.64 g, 66.7 mmol,) was added to 51 mL of concentrated HCl followed by **16** (3.4 g, 22 mmol) and the reaction mixture heated at 90 °C for 1.5 h. TLC (9:1 DCM:MeOH) of an aliquot (pH adjusted to 12 and extracted with EtOAc) showed complete consumption of starting material. The reaction mixture was allowed to cool to room temperature and poured over 475 g of crushed ice. The pH was adjusted to 12 with 7 M NaOH and the reaction mixture extracted with EtOAc (4 x 200 mL). The combined organic layers were dried with $MgSO_4$, filtered, and dried *in vacuo* to afford 2.46 g (70%) of white solid. In most cases the crude product was pure enough to be used directly in the next step; otherwise, flash column chromatography eluting with 9:1 DCM:MeOH gave pure **17**. ^{1}H NMR (400 MHz, DMSO-d_6): δ 7.20 (s, 1H), 5.47 (bs, 2H), 4.63 (bs, 2H), 1.95 (s, 3H). MS (ESI): *m*/*z* 157 (M^+).

4-chloro-7-methyl-1*H*-imidazo[4,5-*c*]pyridine (18)[14] Under argon atmosphere, 20 mL of anhydrous triethylorthoformate (distilled from Na) and 20 mL acetic anhydride (fractionally distilled) was added to 4.0 g (25.4 mmol) of **17** and refluxed for 1 h. The reaction was cooled to room temperature and dried *in vacuo.* The crude compound is boiled in 10% NaOH (10 mL) for 15 minutes and then cooled. Neutralization with HCl precipitates 3.61 g of **18** in 85% yield. ^{1}H NMR (400 MHz, DMSO-d_6): δ 13.3 (bs, 1H) 8.45 (s, 1H), 7.91 (s, 1H), 2.44 (s, 3H). MS (ESI): *m*/*z* 167 (M^+).

7-methyl-1*H*-imidazo[4,5-*c*]pyridin-4-amine (26).[12] To a flask containing **18** (1.50 g, 7.49 mmol) under nitrogen was added 25 mL of anhydrous hydrazine. The mixture was heated to reflux for 30 h. The hydrazine was removed *in vacuo* and co-evaporated with

EtOH (3 x 15 mL). Water (45 mL) and Raney nickel (2.25 g, **CAUTION**: pyrophoric, weighed under tared volume of water) were added and the mixture heated to reflux with vigorous stirring to prevent the Raney Ni from sticking to the stir bar. After 1 h the mixture was filtered, while still hot, through a pad of Celite and the pad washed with hot water (200 mL, CAUTION: prevent leftover Raney Ni from drying). Evaporation of the filtrate *in vacuo* afforded a brown solid. This was re-dissolved in MeOH and silica gel was added. After thorough drying of the mixture the resulting fine powder was loaded onto a flash column and purified by elution with 9:1 DCM:MeOH containing 0.7 M NH_3 to obtain **26** as a white amorphous powder (965 mg, 87%) and starting material (400 mg, 13%). Repetition of this procedure gave yields varying between 60–87%. **26**: R_f = 0.22 (silica gel, 9:1 DCM:MeOH with 0.7 M NH_3); ^{1}H NMR (DMSO-d_6, 400 MHz) δ 12.55 (brs, 1H), 8.08 (s, 1H), 7.42 (s, 1H), 5.86 (brs, 2H), 2.26 (s, 3H) ppm; ^{13}C NMR (DMSO-d_6, 101 MHz) δ 150.79, 140.69, 139.39, 126.57, 108.00, 14.55 ppm; HRMS (DART-TOF) calcd for $C_7H_9N_4H^+$ $[M + H]^+$ 149.0827, found: 149.0834.

General Method for Di-imide Protections

To a flask containing a 6-aminopurine derivative (adenine or **26**, 4.38 mmol) and anhydride (2 eq, 8.77 mmol) was added pyridine (45 mL) and DBU (3 eq, 13.16 mmol). The mixture was heated to reflux for 18 h. Volatiles were removed *in vacuo* and the residue co-evaporated with toluene (3 x 10 mL) to remove residual pyridine. The resulting brown oil was purified by flash column chromatography eluting with 95:5 DCM:MeOH to obtain 1.14 g of white solid. The solid was thoroughly dried *in vacuo* and treated with 10 mL of $SOCl_2$ at reflux for 2 h with a drying tube attached. Volatiles were

removed *in vacuo*; coevaporation with EtOAc (3 x 10 mL) was used to remove residual $SOCl_2$. The resulting off-white solid was then purified by flash column chromatography eluting with 95:5 DCM:MeOH to obtain the di-imide compound as a white amorphous powder.

43: *R*f = 0.61 (silica gel, 9:1 DCM:MeOH); ^{1}H NMR (DMSO-*d*6, 500 MHz) δ 13.88 (brs, 1H), 8.95 (s, 1H), 8.68 (s, 12 1H), 1.28 (s, 12H) ppm; ^{13}C NMR (DMSO-*d*6, 101 MHz) δ 181.18, 152.82, 147.98, 48.60, 21.93 ppm; HRMS (DART-TOF) calcd for $C_{13}H_{15}N_5O_2H^+$ $[M + H]^+$ 274.1297, found: 274.1304.

44: *R*f = 0.29 (silica gel, 95:5 DCM:MeOH); ^{1}H NMR (acetone-*d*6, 400 MHz) δ 8.27 (s, 1H), 8.12 (s, 1H), 2.57 (s, 3H), 1.33 (s, 12H); ^{13}C NMR (acetone-*d*6, 101 MHz) δ 182.63, 145.10, 142.18, 49.06, 22.47, 14.94; HRMS (DART-TOF) calcd for $C_{15}H_{19}N_4O_2H^+$ $[M + H]^+$ 287.1508, found: 287.1508.

General Method for Glycosylation Reactions

To a flask containing the desired heterocycle and NaH (1.5 eq) was added MeCN (final concentration of heterocycle = 0.05 M). The mixture was allowed to stir for 0.5 h followed by addition of α-chlorosugar **3** (1.8 eq). The reaction stirred for 2-8 hr, after which the reaction mixture was filtered through a celite pad. Volatiles were removed *in vacuo* and the resulting crude product was purified by flash chromatography. M_4SI - derivatives can be purified with 1:1 Hex:EtOAc to afford the glycosylated product as a white foam. Products obtained by using Cy_2SI required 2:1 Hex:EtOAc as the eluent for chromatography.

45a: *R*f = 0.33 (silica gel, 97:3 DCM:MeOH); ^{1}H NMR (acetone-*d*6, 400 MHz) δ 8.88 (s, 1H), 8.69 (s, 1H), 8.01 (d, *J* = 8.2 Hz, 2H), 7.92 (d, *J* = 8.1 Hz, 2H), 7.36 (d, *J* = 7.9 Hz, 2H), 7.29 (d, *J* = 8.6 Hz, 2H), 6.79 (dd, *J* = 7.9, 6.1 Hz, 1H), 6.03 – 5.91 (m, 1H), 4.85 – 4.62 (m, 3H), 3.57 (ddd, *J* = 14.4, 7.9, 6.5 Hz, 1H), 2.99 (ddd, *J* = 14.3, 6.1, 2.6 Hz, 1H), 2.42 (s, 3H), 2.39 (s, 3H), 1.35 (s, 12H) ppm; ^{13}C NMR (acetone-*d*6, 101 MHz) δ 181.25, 166.95, 166.77, 154.65, 153.38, 147.36, 146.48, 145.52, 145.16, 132.35, 130.93, 130.80, 130.49, 130.44, 128.49, 128.34, 86.67, 84.06, 76.38, 65.21, 49.29, 37.61, 22.00, 21.96, 21.88, 21.87 ppm; HRMS (ESI-TOF) calcd for $C_{34}H_{35}N_5O_7Na^+$ $[M + Na]^+$ 648.2445, found: 648.2417.

46a: *R*f = 0.35 (silica gel, 97:3 DCM:MeOH); ^{1}H NMR (acetone-*d*6, 600 MHz) δ 8.56 (s, 1H), 8.13 (d, *J* = 0.9 Hz, 1H), 8.05 – 7.96 (m, 2H), 7.93 – 7.83 (m, 2H), 7.40 – 7.33 (m, 2H), 7.33 – 7.24 (m, 2H), 6.90 (dd, *J* = 7.8, 5.8 Hz, 1H), 5.86 (dt, *J* = 6.3, 3.1 Hz, 1H), 4.76 (td, *J* = 4.7, 3.2 Hz, 1H), 4.71 – 4.60 (m, 2H), 3.26 (ddd, *J* = 14.4, 7.8, 6.7 Hz, 1H), 3.09 (ddd, *J* = 14.3, 5.8, 3.0 Hz, 1H), 2.81 (s, 3H), 2.43 (s, 3H), 2.39 (s, 3H), 1.33 (d, *J* = 32.6 Hz, 12H) ppm; ^{13}C NMR (acetone-*d*6, 151 MHz) δ 181.76, 166.50, 166.36, 145.18, 144.78, 143.51, 143.49, 140.30, 138.59, 138.34, 130.55, 130.38, 130.13, 130.11, 27.99, 127.91, 119.95, 86.26, 83.39, 75.55, 64.81, 48.39, 38.59, 21.93, 21.72, 21.61, 21.57, 21.44, 15.51 ppm; HRMS (ESI-TOF) calcd for $C_{36}H_{38}N_4O_7Na^+$ $[M + Na]^+$ 661.2638, found: 661.2629.

References

1) Anan'ev, A.V., Maurin'sh, I. A., Paégle, R. A., Lidak, M. I., Akopian, Z. I. *Bioorg Khim*. **1991**, *17(6),* 855-856.

2) Neyts, J., Balzarini, J., Graciela, A., Chaoyong, Z., Snoeck, R., Zimmermann, A., Mertens, T., Karlsson, A., and De Clercq, E. *Mol. Pharm*. **1997**, *65*, 157.

3) Arico, J. W., Calhoun, A. K., Salandria, K. J., McLaughlin, L. W. *Org. Lett.* **2010***, 12 (1)*, 120-122. Reprinted with permission from *Org Lett.*, 2010, 12 (1), pp 120–122. Copyright 2011 American Chemical Society.

4) Vorbrüggen, H., Ruh-Polenz, C. *Org. React.,* **2000**, *55*, 1.

5) J. A. May and L. B. Townsend. *J. Chem. Soc. Chem. Comm.* **1973**, 64.

6) J. A. May and L. B. Townsend. *J. Chem. Soc. Perkin Trans.* **1975**, *1*, 125.

7) N. Minakawa and A. Matsuda. *Tet. Let.* **1993***, 34 (4),* 661-664.

8) Ogino, M., Yoshimura, Y., Nakazawa, A., Saito, I., Fujimoto, K. *Org. Lett.* **2005**, *7 (14)*, 2853-2856.

9) Z. Kazimierczuk, H. B. Cottam, G. R. Revankar, and R. K. Robins. *J. Am. Chem. Soc.* **1984**, *106*, 6379.

10) H. Seela, H. Rosemeyer, and S. Fischer. *Helv. Chim. Acta*, **1990**, *73*, 1602.

11) P. Serafinowski. *Synthesis,* **1990**, 757.

12) X. Li, D. K. Tuli, D. M. Williams, B. C. Connolly, P. C. Norman, and R. Cosstick. *Nucleic Acids Res.* **1990**, *18*, 4771.

13) Rousseau, R. J.; Robins, R. K. *J. Heterocycl. Chem.* **1965**, *2 (2),* 196.

14) Irani, R. J.; SantaLucia, J. Jr. *Nucleosides, Nucleotides & Nucleic Acids* **2002**, *21*, 737–751.

15) Spencer. Knapp. *Chem. Rev.*, **1995**, *95(6),* 1859–1876.

16) *Bioorg. Med. Chem.* **2006**, *14*, 1935-41.

17) Minghong Zhong, Ireneusz Nowak, and Morris J. Robins* *J. Org. Chem.* **2006**, *71*, 7773-7779

18) Minghong Zhong, Ireneusz Nowak, John F. Cannon, and Morris J. Robins*J. Org. Chem.*, 2006, 71 (11), pp 4216–4221

19) Ireneusz Nowak and Morris J. Robins *J. Org. Chem.*, 2006, 71 (23), pp 8876–8883

20) Roberts, C.; Bandaru, R.; Switzer, C. *Tet. Lett.* **1995**, *36*, 3601-3604.

21) Seela, F.; Gabler, B. *Helv. Chim. Acta* **1994**, *77*, 8322-8323.

ORGANIC LETTERS

2010
Vol. 12, No. 1
120−122

Tetramethylsuccinimide as a Directing/ Protecting Group in Purine Glycosylations

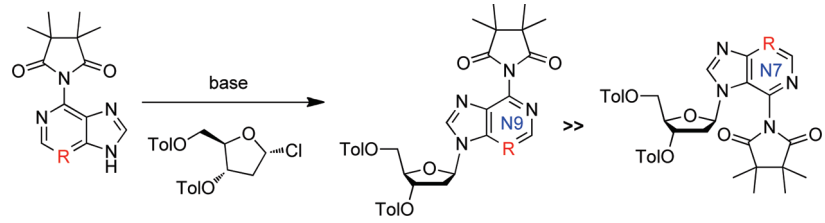
Joseph W. Arico, Amy K. Calhoun, Kerry J. Salandria, and Larry W. McLaughlin*

Department of Chemistry, Merkert Chemistry Center, Boston College, Chestnut Hill, Massachusetts 02467

mclaughl@bc.edu

Received October 29, 2009

ABSTRACT



Tetramethylsuccinic anhydride can be used to protect the exocyclic amine of 6-aminopurine derivatives by forming the corresponding tetramethysuccinimide. X-ray crystallography confirms that the imide carbonyl and the methyl groups are positioned to sterically block the N7 nitrogen so that glycosylations occur with very high regiochemical control at N9. This approach is particularly effective for 3-substituted purines where the substituent tends to block access to N9 and inhibit glycosylation at that site.

Syntheses of purine 2′-deoxynucleosides are challenging, particularly because the glycosylation step must be both regioselective and diasterioselective. Reaction yields of the naturally occurring β-N9 products are often low to moderate. Most preparations of purine 2′-deoxynucleosides employ the sodium salt method,[1] which typically uses a 6-chloropurine heterocycle and 2-deoxy-3,5-di-*O*-(*p*-toluoyl)-α-D-*erythro*-pentofuranosyl chloride.[2,3] For example 6-chloropurine (**1a**, Table 1) reacts to form 57% β-N9 nucleoside and 13% β-N7 with no α-isomers detected.[1] Treatment with NH_3/MeOH deprotects the sugar and displaces the 6-chlorine to form 2′-deoxyadenosine. Small changes in the sterics or electronics of the heterocycle can impact the ratio of coupling products. The favorable coupling ratio for **1a** (N9:N7 = 4.7:1) largely reverses itself for the 6-chloropurine derivative **2a** containing a 3-deaza-3-methyl substituent (N9:N7 = 1.0:2.2).[4] The products of **1a** and **2a** can be separated by flash chromatography, but this is not always the case. Many substrates produce significant amounts of α-nucleoside products, further complicating purification and lowering yields.

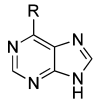
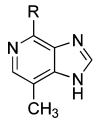
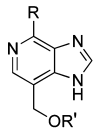
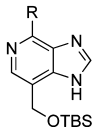
Robins has described the use of purines bearing a 2-alkylimidazole[5,6] or triazole[6] in the 6-position to achieve highly regio- and diastereoselective glycosylations using the sodium salt method. However, the purine coupling partners must be prepared from the corresponding oxopurines or the often expensive chloropurines. In addition, removal of the 2-alkylimidazole was reported to be difficult with most purines.[6] Related studies include the regioselective N9 arylation of purines.[7]

We attempted to use the Robins method in our work with **2a** and other 3-deaza-3-substituted purines, but the 6-(2-alkylimidazolyl) derivatives of **2a** and **3a** could not be prepared, largely because displacement of the 6-chlorine of 3-deazapurines requires harsh conditions (i.e., refluxing

(1) Kazimierczuk, Z.; Cottam, H. B.; Revankar, G. R.; Robins, R. K. *J. Am. Chem. Soc.* **1984**, *106*, 6379–6382.
(2) Rolland, V.; Kotera, M.; Lhomme, J. *Synth. Commun.* **1997**, *27*, 3505–3511.
(3) Hoffer, M. *Chem. Ber.* **1960**, *93*, 2777–2780.
(4) Irani, R. J.; Santalucia, J. *Nucleosides, Nucleotides Nucleic Acids* **2002**, *21*, 737–751.
(5) Zhong, M.; Nowak, I.; Cannon, J. F.; Robins, M. J. *J. Org. Chem.* **2006**, *71*, 4216–4221.
(6) Zhong, M.; Nowak, I.; Robins, M. J. *J. Org. Chem.* **2006**, *71*, 7773–7779.
(7) Bakkestuen, A. K.; Gundersen, L. L. *Tetrahedron Lett.* **2003**, 3359–3362.

10.1021/ol9025028 © 2010 American Chemical Society
Published on Web 12/04/2009

Table 1. Regioselective Purine Glycosylations

heterocycle		β-N9:β-N7	%yield (β-N9)	%yield (β-N7)	%yield (α-N9/α-N7)
	**1a** R = Cl[a]	4.7:1	57	13	0
	1b R = Phthalimide[b]	>99:1	36	0	0
	1c R = M_4SI[c]	>99:1	71	0	6/0
	**2a** R = Cl[d]	1:2.2	39	61	0
	2b R = Phthalimide[b]	1:1.1	15	16	15/16
	2c R = M_4SI	6.4:1	59	9	14/4
	**3a** R = Cl, R' = Ac[b]	1:2.3	26	59	0
	3b R = Phthalimide, R' = Ac[b]	1.2:1	28	23	5/5
	**4a** R = M_4SI	7.6:1	76	11	12/1

[a] See ref 1. [b] Unpublished results. [c] A procedure with KOH and cryptand catalyst TDA-1[4] was used instead of the sodium salt method. [d] See ref 4.

hydrazine followed by Raney nickel reduction[4,8]). Furthermore, the glycosylation products of **3a** could not be separated easily, leading us to seek an alternative procedure.

Here we describe 2,2,3,3-tetramethylsuccinimide (M_4SI) as a new directing/protecting group for the synthesis of β-N9 nucleosides with high regio- and diastereoselectivity.

We reasoned that 6-aminopurines protected with a cyclic imide might perform well as glycosylation partners. The cyclic imide would (i) protect the amino group from side reactions, (ii) permit direct glycosylation to N9, and (iii) be removed by mild base to expose the amino group, with or without deprotection of the sugar, as needed. We first examined phthaloyl protection of adenine using phthalic anhydride. Phthalimide has been reported as a protecting group in DNA synthesis[9,10] and ^{15}N-phthalimide has been used for the introduction of isotopic nitrogen into nucleosides.[11,12] The N^6-phthalimide derivative of adenine could be prepared (**1b**), although the yield was poor (~40%) under all conditions examined. It participated in glycosylation reactions to give almost exclusively the N9 regioisomer but in only 36% isolated yield. Deprotection with NH_3/MeOH removed the phthalimide in less than 15 min and the sugar toluoyl esters after several hours at room temperature. The amine could be reprotected as needed for DNA synthesis, but that approach was less than ideal.

We then considered aliphatic anhydrides that might exhibit better hydrolytic stability, beginning with 1,4-*cis*-cyclohexanedicarboxylic anhydride.[13] Unfortunately this compound polymerizes easily and attempts to prepare simple imides from ammonia or methyl amine have failed. Attempts to prepare the cyclic imide from adenine in this work also failed.

The third compound we considered was tetramethylsuccinic anhydride (M_4SA). Upon formation of the imide with adenine (Figure 1) the methyl groups should partially shield

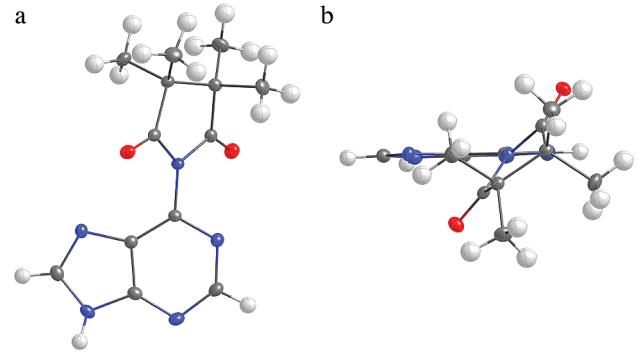
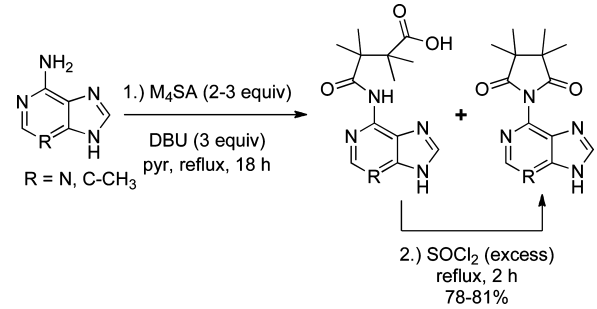


Figure 1. Structure of adenyl-N^6-tetramethylsuccinimide from X-ray diffraction methods: (a) face view and (b) edge view.

the N7 nitrogen from glycosylation. The cyclic anhydride M_4SA is readily obtained in two steps[14] from AIBN.

M_4SA reacted smoothly with adenine and other purines in refluxing pyridine in the presence of DBU (Scheme 1).

Scheme 1. Synthesis of M_4SI-Protected Purines



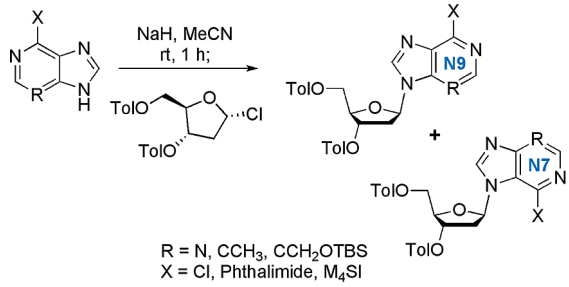
Surprisingly, no reaction took place without base, and other bases such as triethylamine and proton sponge failed to promote the reaction. A small amount of the amide was always present due to incomplete cyclization, but this could be converted to the cyclic imide with refluxing $SOCl_2$, resulting in 78−81% yields of the purine-N^6-tetramethylsuccinimide. It is important to convert the residual amide

(8) Itoh, T.; Yamaguchi, T.; Mizuo, Y. *J. Carbohydr., Nuleosides, Nucleotides* **1981**, *8*, 119–129.
(9) Kume, A.; Sekine, M.; Hata, T. *Tetrahedron Lett.* **1982**, *23*, 4365–4368.
(10) Beier, M.; Pfleiderer, W. *Helv. Chim. Acta* **1999**, *82*, 633–644.
(11) Kamaike, K.; Kinoshita, K.; Niwa, K.; Hirose, K.; Suzuki, K.; Ishido, Y. *Nucleosides, Nucleotides Nucleic Acids* **2001**, *20*, 59–75.
(12) Kamaike, K.; Takahashi, M.; Utsugi, K.; Tomizuka, K.; Ishido, Y. *Tetrahedron Lett.* **1995**, *36*, 91–94.
(13) Hall, H. K., Jr. *J. Am. Chem. Soc.* **1958**, *80*, 6412–6420.
(14) Whitesides, G. M.; Hackett, M.; Brainard, R. L.; Lavalleye, J. P. P. M.; Sowinski, A. F.; Izumi, A. N.; Moore, S. S.; Brown, D. W.; Staudt, E. M. *Organometallics* **1985**, *4*, 1819–1830.

into the cyclic imide since the amide (Scheme 1) does not function well as a blocking group. We obtained the structure of the M_4SI of adenine by X-ray crystallography and two views are shown in Figure 1. In the "face view" (Figure 1a) the four methyl groups can be observed positioned "umbrella-like" above the N7 edge of the adenine ring system. The distance between the imide carbonyl oxygen and the N7 nitrogen is 3.1 Å, and that to C5 is 3.0 Å. These distances likely prevent a coplanar conformation between the imide and heterocycle, and in combination with the four methyl groups seem sufficient to limit access to N7 by the gylcosylation reagent. The "edge view" (Figure 1b) illustrates that the plane of the imide functionality is in fact rotated 52° relative to the plane of the heterocycle, and this rotation is necessary to prevent the steric clash between the imide carbonyls and the N7 nitrogen and C5 carbon.

We next examined a number of adenine-based heterocycles for their ability to undergo glycosylation with M_4SI as a directing/protecting group at the 6-position (Scheme 2) and

Scheme 2. Purine Glycosylations with the Sodium Salt Method



compared these results with those obtained for the corresponding 6-chloropurine, when available (Table 1). Phthalimide-protected heterocycles **1b**, **2b**, and **3b** showed a better directing efffect for N9 than did the 6-Cl derivatives, but only gave low to moderate yields; varying and poorly reproducible amounts of α-nucleosides were always isolated as well. The N7 and N9 regioisomers could be separated but the α and β diasteriosomers could not.

In contrast, glycosylation of M_4SI-protected adenine **1c** yielded only the N9 isomer. The best yield (71%) was obtained with phase transfer conditions using KOH/TDA-1[15] in comparison with NaH (54%). M_4SI performed well with 3-deaza-3-substituted purines. **2c** gave a 6.4:1 ratio of N9 to N7 β-nucleosides, and this ratio increased to 7.59:1 for a 3-deaza analogue bearing a silyl-protected linker (**4a**). Although some coupling reactions generated small amounts of the unwanted regioisomer or diastereomers, with the M_4SI group in place the desired β-N9 products could be resolved from the β-N7 and α-nucleosides by flash chromatography. The α-nucleosides result from anomerization of the chlorosugar (structure illustrated in Scheme 2), which is accelerated in polar solvents.[16] While the M_4SI group is quite effective in blocking access to N7, the 3-substituted purines such as **2c** and **4a** are more hindered nucleophiles; the reaction is then slower to reach completion and larger amounts of α-nucleosides are formed.

The M_4SI group displayed the desired stability that phthalimide lacked. Treatment of the coupling products of **1c** and **2c** with NaOMe/MeOH or 7 M NH_3/MeOH at room temperature removed the two toluoyl esters on the sugar but *not* the M_4SI (64% and 82% yield, respectively), and in 86% yield for the coupling product of **4a** (see the Supporting Information).

Purine nucleosides with substituents at the 3-position are valuable to probe hydration effects in the DNA minor groove. Even a simple methyl substituent at C3 will be directed into the minor groove of duplex DNA where it may disrupt or effect a reorganization of water structure.

As a further example of this protecting group's utility we elaborated M_4SI-protected 2′-deoxy-3-deaza-3-methyladenosine, the coupling product of **2c**, to the DMT-protected phosphoramidite for use in DNA solid-phase synthesis (see the Supporting Information). The M_4SI can be cleanly removed with NH_3/MeOH or concd NH_4OH at 55 °C overnight, both compatible with DNA synthesis protocols. The use of M_4SI-protected phosphoramidites in solid-phase DNA synthesis will be reported elsewhere.

M_4SI is a valuable new directing/protecting group that allows rapid access to β-N9 6-aminopurine-2′-deoxynucleosides and functions as an effective base-labile protecting group. It is especially valuable for 3-substituted purines. It should also serve in other applications where bidentate amine protection is desired.

Acknowledgment. We thank Dr. Bo Li for the crystallographic analysis of **1c** and Dr. Nick Greco for critical discussions. This work was supported by an award from the NSF to L.W.M. (MCB 0451488).

Supporting Information Available: Experimental procedures and characterization of all compounds along with NMR spectra. This material is available free of charge via the Internet at http://pubs.acs.org.

OL9025028

(15) Seela, F.; Roseyer, H.; S., F. *Helv. Chim. Acta* **1990**, *73*, 1602–1611.

(16) Kotick, M. P.; Szantay, C.; Bardos, T. J. *J. Org. Chem.* **1969**, *34*, 3806–3811.

Chapter 3: Use of Tetramethylsuccinimide to Study Effects of Small Molecule Binding in the Minor Groove

3.1 Chemical Synthesis of Modified DNA Oligomers

Our primary goal involved tethering hydroxyl and ammonium functionality off the 3-position of dA. Synthetic access to minor groove modifications is a requirement in order to study how small molecules, such as water and ammonium, bind and associate within the minor groove of DNA (**Figure 1**). Somewhat strenuous conditions (*i. e.* high temperature) are required to functionalize the benzyl position of 3-deaza-3-methylpurine. Nucleosides are more sensitive to these types of conditions and have tendencies to decompose through depurination at higher temperatures compared to their

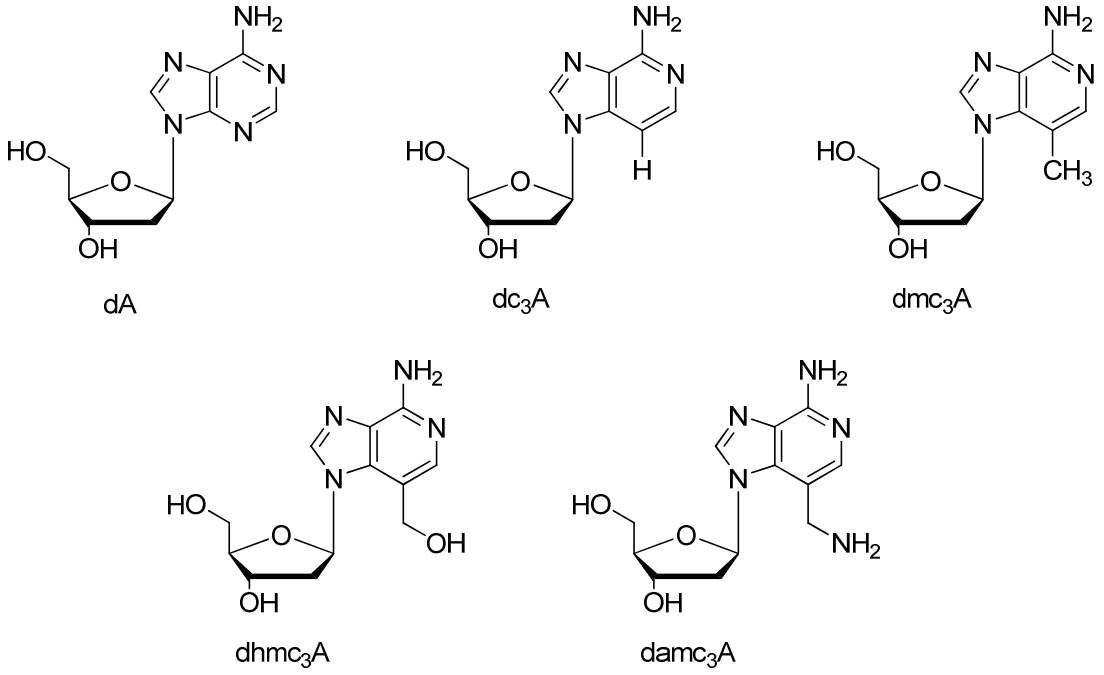
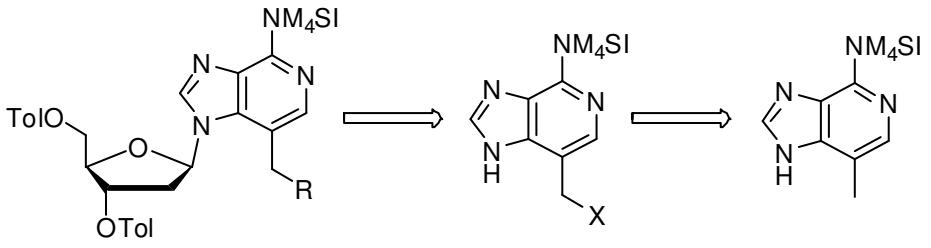


Figure 1. Modifications of dA used to study small molecule binding in the minor groove

nucleobase predecessors. Therefore, it is more desirable to functionalize the heterocycle prior to glycosylation which implies the steric size of the 3-position will only further increase (**Scheme 1**). Tetramethylsuccinimide has served as a valuable tool to direct the glycosylation reaction to desired β-N9 products in the presence of steric hinderance at the 3-position. Using this protecting group will allow us to synthesize minor groove modified nucleosides in more acceptable yields. The modified nucleosides can then be deprotected and further functionalized to the corresponding phosphoramidites.

Scheme 1. Retrosynthesis of 3-substituted 2'-deoxyadenosine Analogues.



Phosphoramidites are the standard building block of chemical, solid-phase DNA synthesis.[1] Solid-phase DNA synthesis elongates in a 3' to 5- direction (**Figure 2**). The 5'-hydroxyl on the sugar is protected with dimethoxyltrityl. This group is highly sensitive to acid and can be removed using 3% dichloroacetic acid in dichloromethane. Attached to the 3'-hydroxyl is a phosphite group functionalized with diisopropylamino and 2-cyanoethoxy groups. Diisopropylamino groups can be activated within the course of the synthesis with tetrazole for displacement by incoming 5'-deprotected phosphoramidites. The new linkage is then oxidized with iodine and water in pyridine to prevent further reactivity of

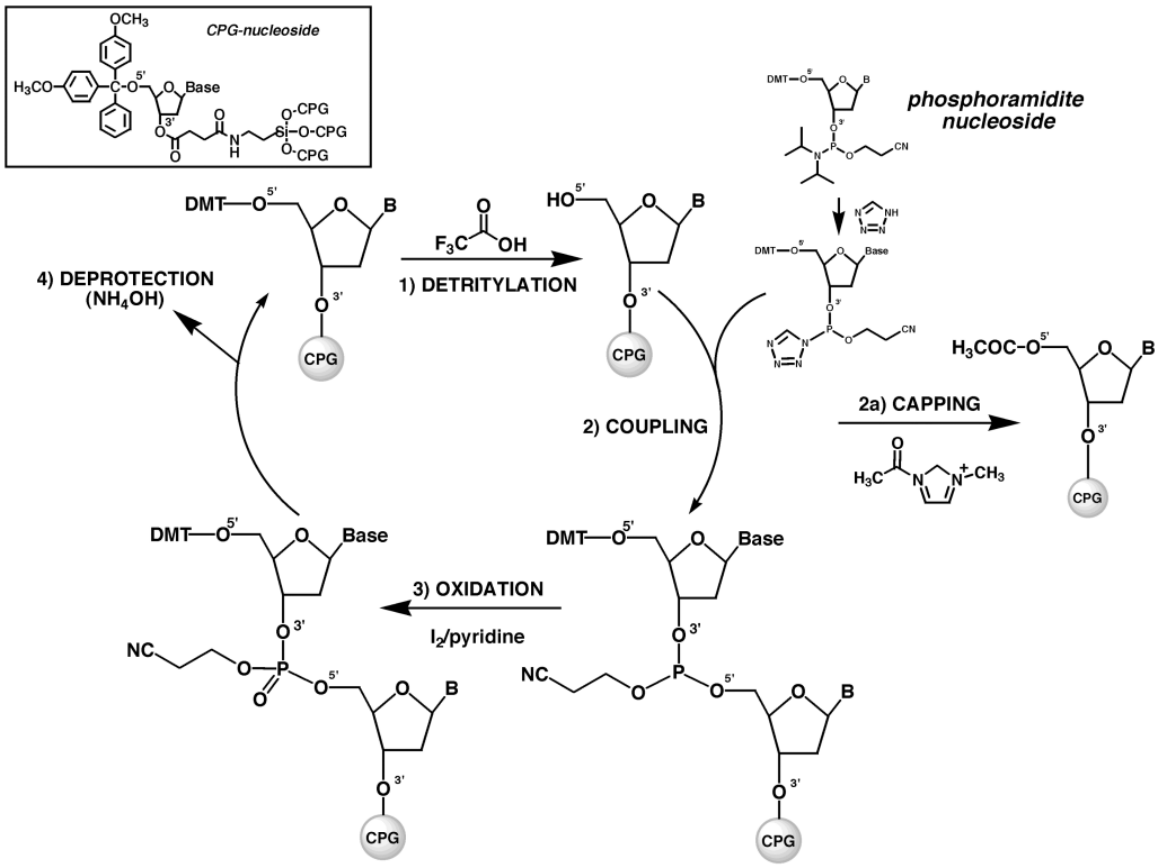


Figure 2. The DNA solid-phase synthesis cycle using the phosphoramidite method. Source: Wikimedia Commons.

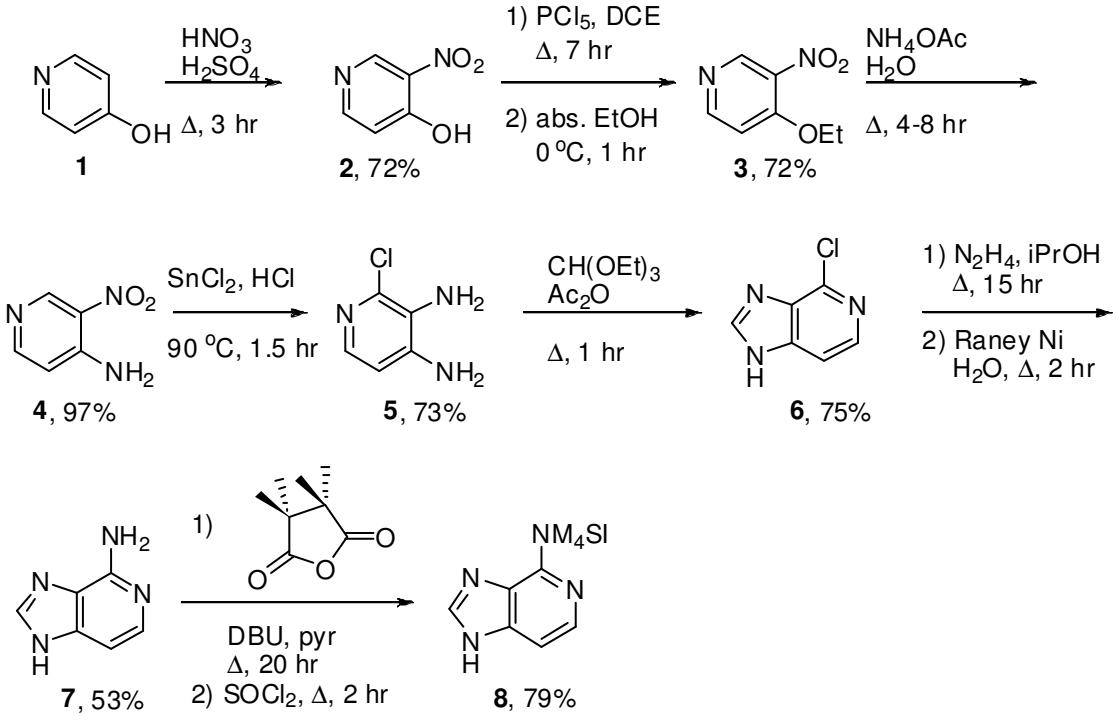
the phosphite during synthesis. The 2-cyanoethoxy group and heterocycle protecting groups can be deprotected post-synthesis with ammonium hydroxide at 55 ^{o}C.

The oligomer used in these studies contains a central core of 4 dA residues flanked by G:C base pairs. This sequence was designed to exhibit small molecule binding operties of A-tract DNA. The non-self complementary nature of this sequence allows the effects of a single or multiple modifications to be measured.

3.2 Synthesis of dc_3A, dmc_3A, $dhmc_3A$, and $damc_3A$ Phosphoramidites

We began with elaboration of dc_3A. The use of M_4SI- as a protecting/directing group eliminated the need for the chloro-substitution at the 2-position of the heterocycle, so a new synthesis to obtain the c_3A heterocycle was designed. Using 4-hydroxypyridine (**1**) as the starting material, M_4SI-protected 3-deazaadenine **8** was synthesized in 7 steps (**Scheme 2**). The synthesis commences with nitration of **1** to install the N7 nitrogen of c_3A. Conversion of the hydroxypyridine **2** to ethyl ether **3** is accomplished by first treating **2** with PCl_5 followed by absolute ethanol. This activates the 4-position towards displacement with NH_4OAc to install the N9 nitrogen (**4**). The following steps are analogous to the synthesis of mc_3A as shown in Section 2.2. Treatment with $SnCl_2$ in HCl reduces the nitro group and installs chloro-functionality at the 6-position to give **5**. The

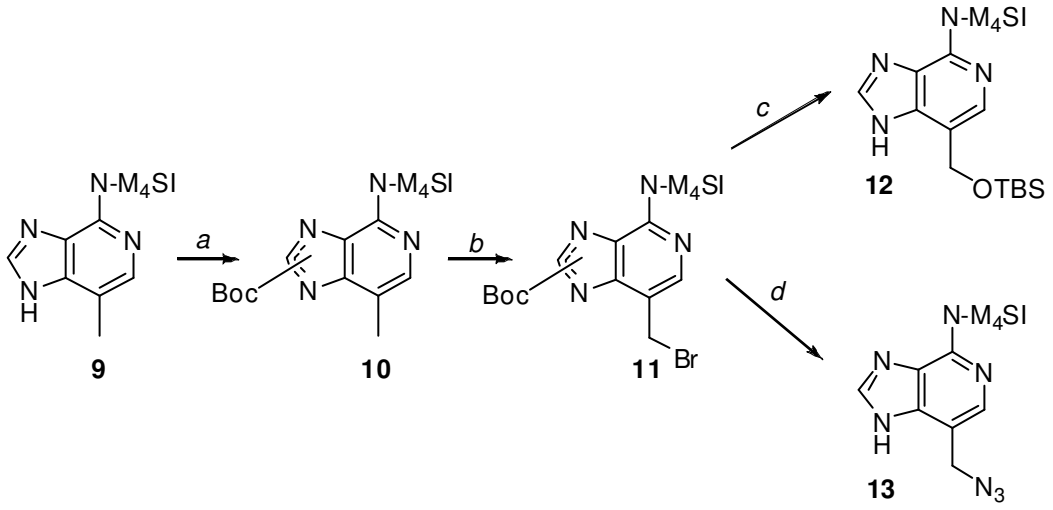
Scheme 2. Sythesis of M_4SI-protected 3-deazaadenine (**8**).



imidazole ring is closed using triethylorthoformate and acetic anhydride to give **6**, followed by displacement of the chlorine with hydrazine and Raney Nickel reduction yields 53% of c_3A heterocycle **7**. The exocyclic amine can be treated with tetramethylsuccinyl anhydride in the presence of DBU to produce the functionalized heterocycle for glycosylation (**8**).

To further functionalize the 3-position, 3-deaza-3-methylpurine undergoes benzyllic bromination after t-Boc protection of the heterocycle. This intermediate is a branching point for further elaboration of the 3-position (**Scheme 3**). The tethered hydroxyl is first installed by displacement of the bromide substituted heterocycle with sodium acetate in DMF at slightly elevated temperatures. The elevation in temperature in a polar solvent is enough to cleave the Boc group from the imidazole moiety, which is surprising since this protecting group is most commonly removed with 1:1 TFA:CH_2Cl_2.

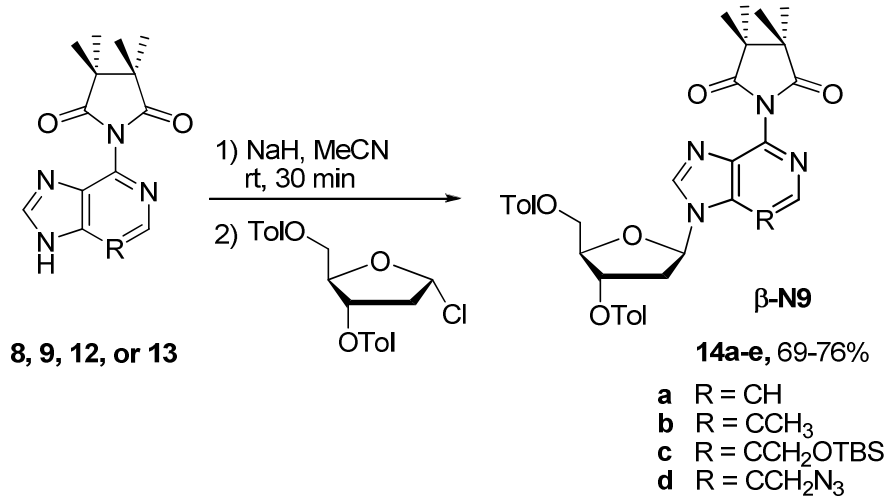
Scheme 3. Functionalization of M_4SI-mc_3A to Install Tethered Small Molecules.



a) Boc_2O, DMAP, CH_2Cl_2, rt, 30 min, >95%. *b*) NBS, $(BzO)_2$, CCl_4, 80 ^oC, 4 hr, 80%. *c*) NaOAc, DMF, 60 ^oC, 3 hr then 7N NH_3 in MeOH, 2 hr, then TBSCl, DIPEA, 4:1 CH_2Cl_2:DMF, rt, 20 hr, 75%. *d*) NaN_3, DMF, 45 ^oC, 90 min, 95+%.

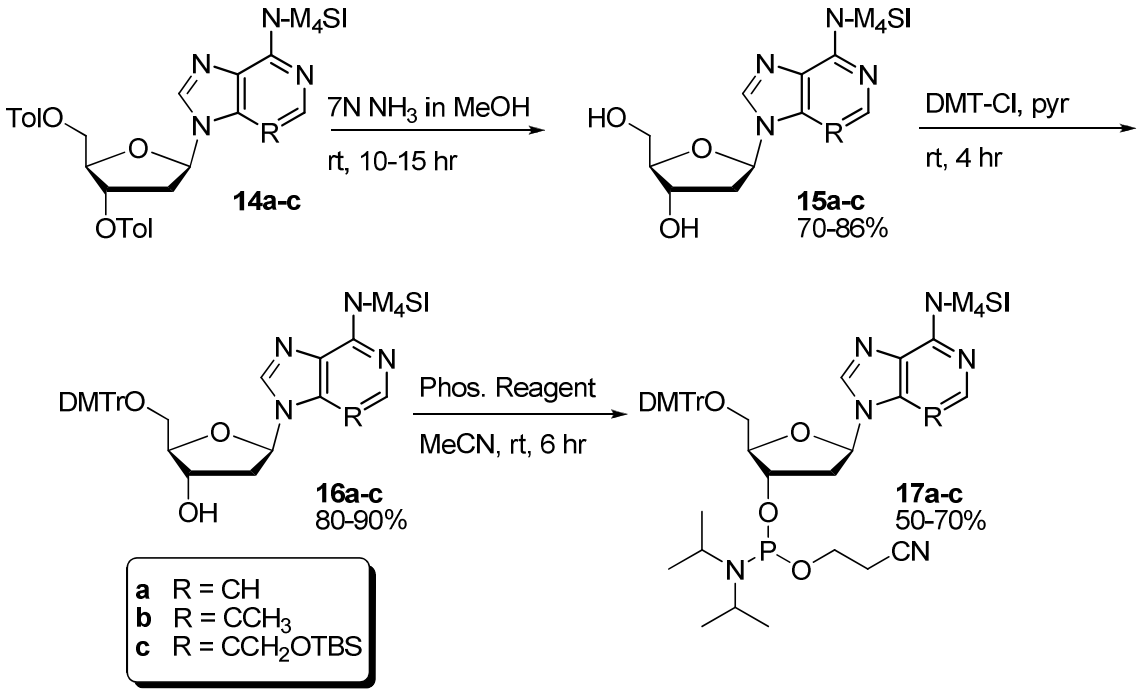
The free hydroxyl results from acetate deprotection with ammonia in methanol. The hydroxyl must be protected prior to glycosylation to prevent side reactions with sodium hydride. Amine functionality is first installed by displacement of the bromide with sodium azide at elevated temperatures. This serves to prevent the free amine from side reactions during glycosylation. The azide can be reduced and selectively protected after glycosylation.

Scheme 4. Glycosylation of 3-Modified Adenine Analogues Using M_4SI- as a directing group



The four M_4SI- protected heterocycles (**8, 9, 12,** and **13**) are next glycosylated to provide the modified nucleosides (**Scheme 4**). The c_3A-protected nucleoside produces β-N9 isomer exclusively, much like M_4SI-protected adenine, due to the absence of steric bulk at the 3-position. Much to our satisfaction, addition of bulkier substitutions off the 3-position did little to shift the ratio of glycosylation products. This proves the ability of M_4SI- to direct the glycosylation and yield β-N9 products in approximately 70% yield, as opposed to the 20-50% glycosylation yields that were considered acceptable before our work.

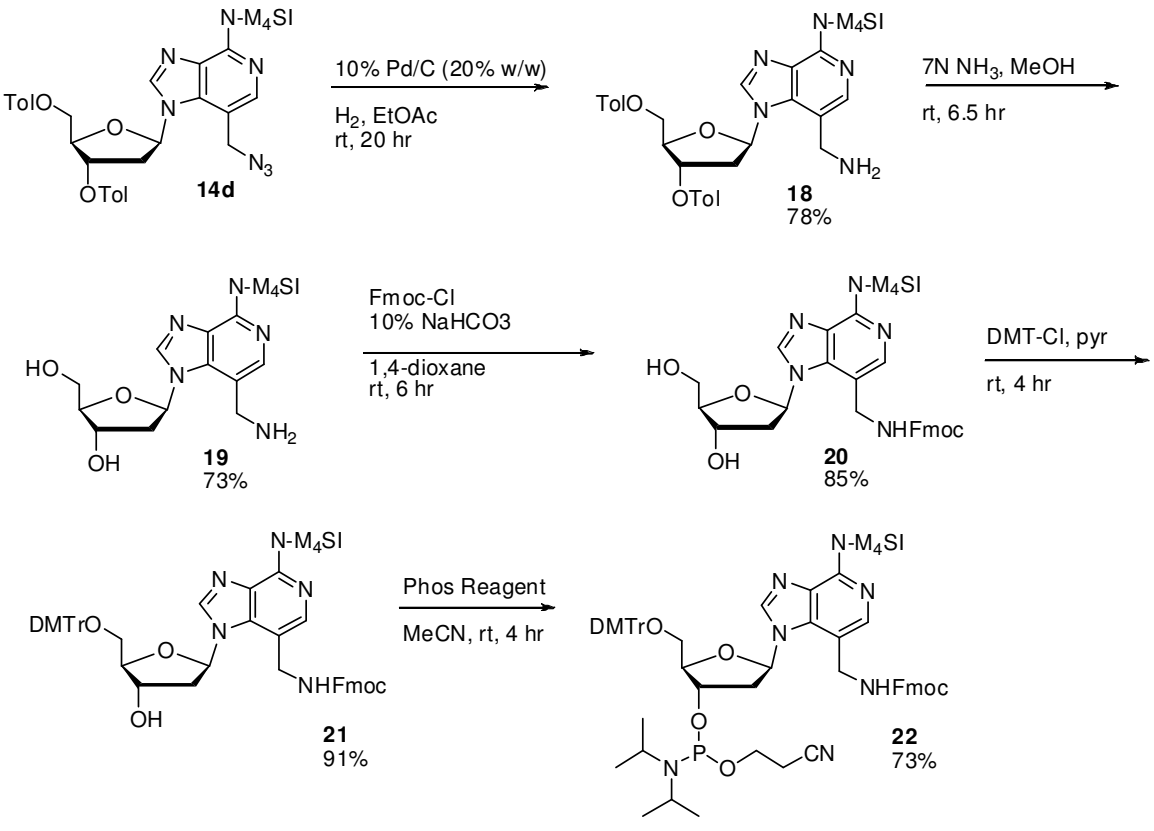
Scheme 5. Synthesis of dc3A (**17a**), dmc3A (**17b**), and dhmc3A (**17c**) Phosphoramidites



Once the protected nucleosides of dc3A (**14a**), dmc3A (**14b**), and dhmc3A (**14c**) are isolated, steps to the phosphoramidites **18a-c** are straight-forward (**Scheme 5**). First, the sugar is deprotected with 7 N ammonia in methanol over 10 hr. Then, the 5'-hydroxyl can be selectively protected with dimethoxytrityl ether functionality. The advantage of using dimethoxytrityl as a protecting group is two-fold: the tertiary substituted, large protecting group selectively protects primary hydroxyls over secondary hydroxyls and it is extremely labile towards acid leading to complete and effective deprotection during oligomer synthesis. Lastly, the phosphite group is installed on the 3'-hydroxyl. Two common reagents for this step are 2-cyanoethyl *N,N,N,N*-tetraisopropyl phosphane and chloro-2-cyanoethyl-*N,N*-diisoproylphosphane. They can be used interchangeably as they simply differ in how they are activated.

The aminomethyl ($damc_3A$) phosphoramidite requires additional steps to completely functionalize and protect it for DNA synthesis (**Scheme 6**). After glycosylation, the azide is reduced with 10% palladium on charcoal (20% w/w) and hydrogen to afford free amine **18**. Deprotection of the toluoyl results in M_4SI-protected $damc_3A$ **19**. The amine can be selectively protected over the hydroxyl functionality using Fmoc-Cl in dioxane with 10% $NaHCO_3$ (**20**). Next, the 5'-hydroxyl is protected to give the DMT-protected nucleoside **21**. To complete the synthesis of damc3A phosphoramidite **22**, **21** is phosphitylated using standard conditions.

Scheme 6. Synthesis of $damc_3A$ Phosphoramidite **22**.



3.3 Impact of Structural Water Mimic in the Minor Groove

Using the modified phosphoramidites, DNA of varying sequence and functionality can be synthesized. We prepared four types of oligonucleotides: (i) those containing 3-deazaadenine, (ii) those containing 3-deaza-3-methyladenine, (iii) those containing 3-deaza-3-hydroxymethyladenine, and (iv) those containing 3-deaza-3-aminomethyladenine. The design parameters were as follows: The first modified oligos did not contain the N3 nitrogen, a site that is used in the formation of the spine of hydration but otherwise introduces no unfavorable steric effects. In the methyl-containing sequences, the N3 nitrogen was also removed, and additional unfavorable steric effects were introduced into the minor groove through the presence of one or more methyl groups. Third, hydroxyls were introduced onto the methyls to generate a covalently bound water mimic capable of interacting by hydrogen bonding in the minor groove. Lastly, amines were covalently tethered onto the methyl groups to serve as both an amine hydrogen bond donor or ammonium cation depending on the pH of the solution.

Thermodynamic parameters were obtained for the formation of all duplexes using their melting curves at 5 uM duplex concentration in 20 mM phosphate buffer, pH 7 and 1 M NaCl (**Table 1**). Even the simple elimination of a single N3 nitrogen from a centrally located dA residue (duplex **2**) resulted in a 2.4 °C decrease in the melting temperature (T_M), consistent with previous reports. In that same study, the introduction of three dA residues lacking the N3 nitrogen resulted in a 12 °C change in T_M, which also compares favorably with the 12.3 °C reduction in T_M observed here with four deazaadenines

Table 1. Thermodynamic Stabilities of Duplexes Containing dA Analogues

Entry	Core Sequence d(CCGG XXXX CGCC)	T_m †	ΔH kcal/mol	ΔS Cal/molK	ΔG‡ kcal/mol
1	AAAA	63.4	-95.1 (± 4.8)	-255 (± 15)	-19.1 (± 0.6)
2	AA$\boldsymbol{A}^{H}$A	61.0	-82.0 (± 5.8)	-220 (± 18)	-16.4 (± 0.4)
3	AA$\boldsymbol{A}^{CH3}$A	58.3	-70.8 (± 3.7)	-188 (± 12)	-14.8 (± 0.3)
4	AA$\boldsymbol{A}^{CH2OH}$A	60.9	-71.7 (± 3.9)	-185 (± 11)	-16.4 (± 2.3)
5	AA$\boldsymbol{A}^{CH2NH3+}$A	62.9	-73.1 (± 1.6)	-191 (± 4.6)	-17.5 (± 1.1)
6	$(\boldsymbol{A}^{H})_4$	51.1	-76.1 (± 2.9)	-209 (± 9.0)	-13.9 (± 0.2)
7	$(\boldsymbol{A}^{CH3})_4$	44.4	-51.7 (± 4.1)	-137 (± 13)	-10.8 (± 0.4)
8	$(\boldsymbol{A}^{CH2OH})_4$	54.7	-82.0 (± 5.5)	-224 (± 16)	-15.2 (± 0.7)
9	$(\boldsymbol{A}^{CH2NH3+})_4$	44.0	-50.1 (± 6.9)	-131 (± 22)	-11.1 (± 0.4)
9*	$(\boldsymbol{A}^{CH2NH2})_4$, pH 9	50.4	-38.6 (± 1.2)	-93.2 (3.2)	-10.8 (± 0.3)
10	$\boldsymbol{A}^{CH2NH3+}$A$\boldsymbol{A}^{CH2NH3+}$A	61.1	-78.6 (± 2.9)	-212 (± 7.0)	-15.6 (± 1.2)

†Determined at 5 uM duplex concentration in 20 mM phosphate buffer, pH 7 (unless otherwise noted) and 1M NaCl. ‡ Calculated at 298 K.

present (duplex **6**). We then examined the methyl-substituted sequences (duplexes **3** and **7**). The 3-deaza-3-methyl analogues both lack the N3 nitrogen but also add steric bulk within the minor groove. Introduction of a single methyl group (duplex **3**) reduced the T_M value by 2.7 °C relative to a sequence containing a single 3-deaza analogue (duplex **2**) and by 5.1 °C relative to the unmodified standard. The presence of four analogues containing methyls (duplex **7**) reduced the T_M by a dramatic 19.0 °C, attesting to the

duplex instability resulting from increased steric effects in the minor groove. The hydroxymethyl and aminomethyl groups are sterically even larger than a methyl group and could have an even more dramatic destabilizing effect should these groups be unable to take part in advantageous interactions in the minor groove. Introduction of one $-CH_2OH$ group did result in helix destabilization relative to the control sequence (compare the T_M for duplex **4** with that for duplex **1**). However, relative to the single-methyl sequence (duplex **3**), the T_M for the hydroxymethyl sequence was some 2.6 °C higher. This observation is further supported by the sequence with four $-CH_2OH$ groups (duplex **7**), which exhibited a T_M 10.3 °C higher than that for the sequence containing four methyl groups. The addition of one $-CH_2NH_3^+$ (duplex **5**) resulted in greater stabilization relative to methyl and hydroxymethyl (duplex **3** and **4**, respectively). The single tethering of the ammonium offers near native stability. Direct comparison of the two alpha plots shows that the T_M values of the two complexes are within error (**Figure 3**).

Substitution with 4 aminomethyl derivatives (duplex **9**) at pH 7 was similar in respect to the 4 methyl substitutions (duplex **7**) differing by 0.4 ^oC. Four consecutive amines function to stabilize the helix more than the ammoniums. Duplex **9*** has a T_M of 50.4 ^oC, which appears more similar to the hydroxymethyl substituted duplex **7**, suggesting it is acting as a hydrogen bond donor, albeit a somewhat weaker donation than that of the hydroxyl groups. We concluded that 4 consecutive ammonium cations (duplex **9**) results in charge-charge repulsion within the minor groove. Duplex **10** was designed to indicate whether the presence of multiple groups leads to a more stable duplex than

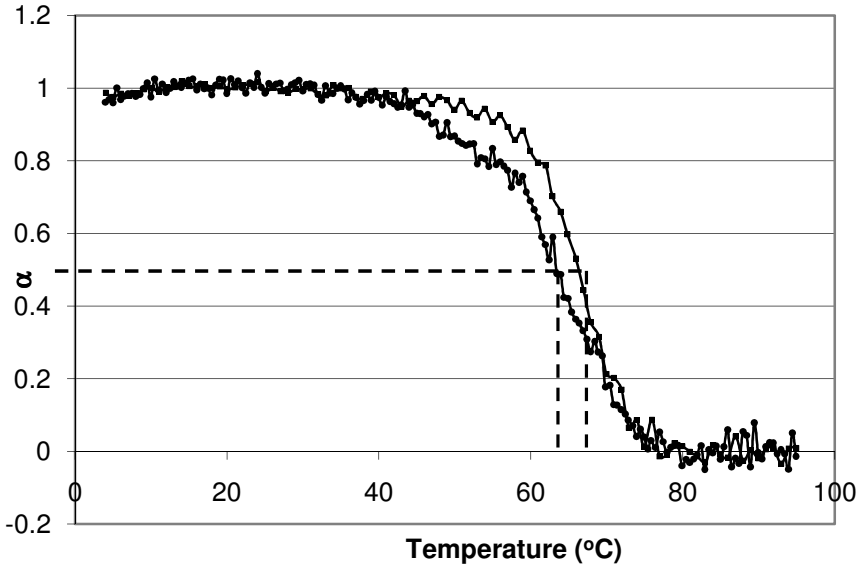


Figure 3. Alpha plots comparing the native sequence, d(GGCC AAAA CGCC) (squares), and one amc_3A substitution, d(GGCC AA$\boldsymbol{A}^{CH2NH3+}$A CGCC) (circles). The melting curve of the native is 63.4 °C and that of the amino-substituted oligo is 62.9 °C.

native, since a single substitution has a stabilizing effect while consecutive substitutions lead to an overall destabilization of the helix. No significant additional cooperativity appeared to be present for any analogue duplex.

Next, we investigated how the overall structure of the duplex is affected by incorporation of these analogues by CD spectroscopy (**Figure 4**). All of the spectra suggest the presence of B-form helices. The CD spectra for the native duplex **1** and duplex **6** containing four dc_3A residues are the most alike of the four spectra, suggesting similar chromophoric stacking (dA vs dc_3A) within the helix. The interactions of the purine heterocycles are dramatically different for duplex **7**, in which four methyl groups have been added to the central core of base pairs. This spectrum suggests a quite different stacking of the dmc_3A chromophores, leading to dramatic shifts in the magnitude of the

positive and negative displacements in the observed CD spectrum. The spectrum for duplex **8** (4 $dhmc_3A$) is intermediate in character relative to those of duplexes **6** and **7**. It seems likely that the minor-groove interactions present with the structural water mimic drive the conformation back toward a more native-like B-form, but much like the temperature data, the final native B-like structure is not fully attained. Lastly, the spectrum of duplex **9** shows complete disturbance of B-form DNA. The spectrum directly relates to the temperature data for duplex 9, further supporting that sequential cations in the minor groove do not offer stability or structural integrity to the helix.

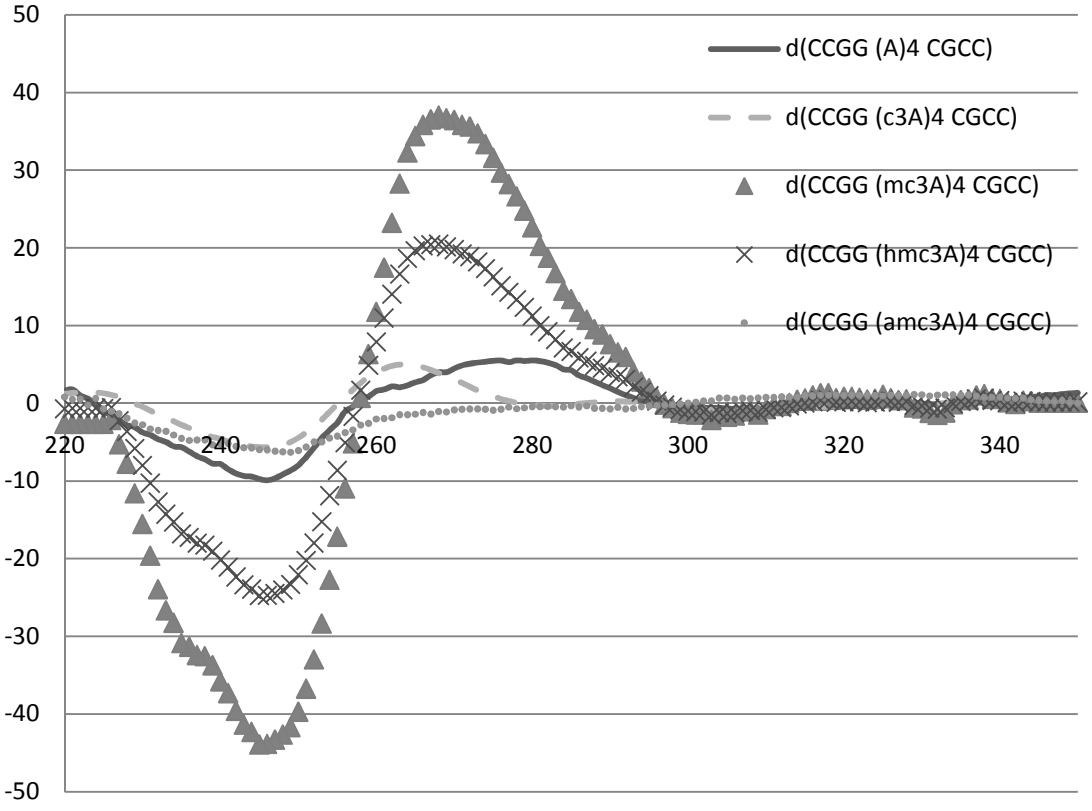


Figure 4. CD spectra (ellipticity versus wavelength) at 3 uM concentration of duplexes **1, 6, 7, 8, and 9**

3.4 Conclusions

Use of tetramethylsuccinimide as a directing and protecting group allows synthetic access of minor groove modified nucleosides in higher yields than their corresponding chloro-derivatives. These modified nucleosides can be elaborated to their corresponding phosphamidites and used for chemical synthesis of DNA. Deletion of the N3 nitrogen on adenine (dc_3A) destabilizes the helix, but not as much as incorporation of steric bulk resulting from incorporation of dmc_3A. Tethering functional groups that mimic water ($dhmc_3A$) or ammonium ($damc_3A$) binding in the minor groove increases the steric size of these minor groove substitutions, so the overall effect is still destabilizing relative to the native sequence. However, stabilization relative to sequences containing dmc_3A support that small molecule binding in the minor groove is important to the overall stability of duplex DNA.

Experimental

4-Hydroxy-3-nitropyridine (2).[2] Fuming nitric acid (58 mL, >99.5%) was slowly added to fuming sulfuric acid (49 mL, 26-29.5% free SO_3) at 0 °C. This was followed by the addition of 4-hydroxypyridine (20 g, 210 mmol) portion wise over 15 min. The reaction was slowly heated until the exothermic evolution of N_2O_4 occurred. The heat source was removed and after the reaction slowed, the mixture was heated to 120 °C for 1 h. The reaction mixture was cooled to ambient temperature and poured slowly over ice. The mixture was neutralized with conc. NH_4OH and chilled overnight. Yellow crystals were collected by filtration and washed with a small amount of ice water to yield **24** (21.3 g, 72%). ^{1}H NMR (400 MHz, DMSO-*d*6) δ 8.78 (s, 1H); 8.52 (s, 1H); 7.76 (d, *J* = 7.6 Hz 1H); 6.46 (d, *J* = 7.6 Hz, 1H). HRMS (ESI-TOF): calcd for $C_5H_5N_2O_3$: 141.03002, found: 141.02987.

4-Ethoxy-3-nitropyridine (3)[3]**.** To PCl_5 (50.1 g, 237 mmol) dissolved in dichloroethane (180 mL) was added 4-hydroxy-3-nitropyridine (27.7 g, 197 mmol) and the solution was refluxed for 7 h. After the reaction was complete (verified by TLC, 7:3:1 EtOAc:MeOH:28% aq. NH_3), the mixture was cooled to 0 °C and 111 mL of absolute ethanol was added slowly via a syringe with vigorous stirring. A thick, white slurry formed with an increase in temperature. The slurry was stirred at 0 °C for 20 min and then refluxed for 1 h. The flask was cooled at 4 °C for 2–3 h, filtered and washed with cold ethanol to yield white crystals of **20** as the HCl salt (32.24 g, 80% (87% reported)) after drying *in vacuo*. ^{1}H NMR (400 MHz, DMSO-*d*6): δ 9.10 (s, 1H); 8.91 (bs, 1H); 8.74

(d, *J* = 6.4 Hz, 1H); 7.56 (d, *J* = 6.4 Hz, 1H); 4.39 (q, *J* = 6.8 Hz, 2H); 1.38 (t, *J* = 7.2 Hz, 3H). HRMS (ESI-TOF): calcd for $C_7H_9N_2O_3$: 169.06132, found: 169.06137.

4-Amino-3-nitropyridine (4)[3] Ammonium acetate (14.7 g, 191 mmol) was dissolved in water (24 mL) and 4-ethoxy-3-nitropyridine hydrochloride (9.81 g, 47.9 mmol) was added. The solution was refluxed for 4–8 h. The reaction was monitored by TLC (10:1 EtOAc:triethylamine) and additional ammonium acetate was added if necessary. When the reaction was complete, the mixture was cooled to ambient temperature and adjusted to pH 8 with NH_4OH and stirred for 1 h at 0 °C. The resulting precipitate was filtered and dried overnight to yield **19** as a yellow powder (6.58 g, 97%). ^{1}H NMR (400 MHz, DMSO-*d*6) δ 8.97 (s, 1H); 8.13 (d, *J* = 6.0 Hz, 1H); 7.93 (bs, 2H); 6.89 (d, *J* = 6.0 Hz, 1H). HRMS (ESI-TOF): calcd for $C_5H_6N_3O_2$: 140.04500, found: 140.04663.

3,4-diamino-2-chloropyridine (5)[4,5] To 25 mL of conc. HCl was added $SnCl_2$ (33.1 g, 175 mmol) followed by 4-amino-3-nitropyridine (5.0 g, 35.7 mmol) added portion wise to the solution over 10 minutes. The reaction was then refluxed for 1.5 h, after which TLC (9:1 DCM:MeOH) of a neutralized aliquot extracted with EtOAc showed complete conversion. The reaction was then cooled to ambient temperature and poured over 400 mL of crushed ice. The solution was neutralized with 7 M NaOH and diluted with 200 mL of water to dissolve most of the thick white precipitate that formed. Additional NaOH was added to maintain the pH at ~12. The slurry was extracted with ethyl acetate (4 x 100 mL), dried over $MgSO_4$, and evaporated *in vacuo* to yield **17** as a white to rose-colored solid that was used without further purification (3.78 g, 73%). ^{1}H NMR (400 MHz,

DMSO-*d*6): δ 7.28 (d, *J* = 5.2 Hz, 1H); 6.43 (d, *J* = 5.2 Hz, 1H); 5.76 (s, 2H); 4.66 (s, 2H). HRMS (ESI-TOF): calcd for $C_5H_7ClN_3$: 144.03285, found: 144.03290.

4-chloro-1*H*-imidazo[4,5-*c*]pyridine (6)[6] Under inert atmosphere, triethylorthoformate (33.3 mL, freshly distilled over sodium) and acetic anhydride (33.3 mL, fractionally distilled) were added to a flask containing 3,4-diamino-2-chloropyridine (6.67 g, 46 mmol). The reaction was then heated at reflux with stirring for 1 h, cooled to ambient temperature, and concentrated *in vacuo*. The resulting yellow solid was dissolved in minimal 10% NaOH and heated to 95 °C for 10 minutes. The solution was then neutralized to pH 6 with acetic acid and cooled at 4 °C overnight. White crystals of **15** (5.33 g, 75%) were collected by filtration, washed with cold water and dried *in vacuo*. ^{1}H NMR (400 MHz, DMSO-*d*6): δ 8.44 (s, 1H); 8.01 (d, *J* = 5.6 Hz, 1H); 7.60 (d, *J* = 5.6, 1H). HRMS (ESI-TOF): calcd for $C_6H_5ClN_3$: 154.01720, found: 154.01638.

1*H*-imidazo[4,5-*c*]pyridin-4-amine (7)[6] To a flask containing **6** (200 mg, 1.31 mmol) was added 3.53 mL of anhydrous hydrazine and 2.35 mL of isopropanol. The mixture was heated to reflux for 14 h (anhydrous conditions not needed). The reaction was concentrated *in vacuo*. Water (7 mL) and Raney nickel (329 mg, **CAUTION**: pyrophoric, weighed under tared volume of water) were added and the mixture heated to reflux with vigorous stirring. After 1 h the mixture was filtered, while still hot, through a pad of Celite and the pad washed with hot water (200 mL, CAUTION: prevent leftover Raney Ni from drying). Evaporation of the filtrate *in vacuo* afforded a tan solid. The crude compound was purified via solid pack with flash column chromatography (10% MeOH and 0.7 N NH_3 in DCM) to obtain **7** as a white amorphous powder (93 mg, 53%). ^{1}H

NMR (400 MHz, methanol-d_4): δ 7.99 (s, 1H), 7.54 (d, J = 9.2 Hz, 1H), 6.76 (d, J = 5.6 Hz, 1H), 5.00 (bs, 2H). HRMS (ESI+): calc. 135.06707 found. 135.0651.

***N*-(1*H*-imidazo[4,5-*c*]pyridin-4-yl)-2,2,3,3-tetramethylsuccinimide (8).**[7] To a flask containing **7** (93 mg, 0.693 mmol) and 2,2,3,3-tetramethylsuccinic anhydride (216 mg, 1.39 mmol) was added pyridine (6.4 mL) and DBU (0.311 mL, 2.08 mmol). The mixture was heated to reflux until TLC indicated complete conversion of starting material (20 hr). Volatiles were removed *in vacuo.* The resulting brown oil was purified by flash column chromatography eluting with 95:5 DCM:MeOH to obtain 154 mg (82%) of white solid. The solid was thoroughly dried *in vacuo* and treated with 1.5 mL of $SOCl_2$ at reflux for 2 hr. Volatiles were removed *in vacuo*; coevaporation with EtOAc (3 x 10 mL) was used to remove residual $SOCl_2$. The resulting off-white solid was then purified by flash column chromatography eluting with 95:5 DCM:MeOH to obtain 100 mg (53%) of **8**. ^{1}H NMR (400 MHz, MeOH-d_4): δ 8.39 (s, 1H), 8.34 (d, J = 5.2 Hz, 1H), 7.73 (d, J = 5.6 Hz, 1H), 1.381 (s, 12H) ppm. ^{13}C NMR (100 MHz, MeOH-d_4): δ 181.79, 144.91, 142.62, 140.43, 20.44, 19.57 ppm. HRMS (ESI+): calc. 273.13515 found. 273.13573.

***tert*-butyl 7-methyl-4-(3,3,4,4-tetramethyl-2,5-dioxopyrrolidin-1-yl)-1*H*-imidazo[4,5-*c*]pyridine-1-carboxylate** and ***tert*-butyl 7-methyl-4-(3,3,4,4-tetramethyl-2,5-dioxopyrrolidin-1-yl)-3*H*-imidazo[4,5-*c*]pyridine-3-carboxylate (10)**[7] To a flask containing **9** (2.85 g, 9.95 mmol) was added Boc_2O (3.04 g, 13.9 mmol) and DMAP (243 mg, 1.99 mmol). DCM (52 mL) was added and the mixture stirred at rt for 30 min (anhydrous conditions not needed). The starting material gradually dissolved to produce a yellow homogeneous solution. TLC (97:3 DCM: MeOH) indicated that a small amount

of starting material remained. More Boc_2O (1.1 g, 5 mmol) and DMAP (122 mg, 1 mmol) were added to drive the reaction to completion. After allowing the mixture to stir for another 15 min, TLC indicated that no starting material remained. Volatiles were removed *in vacuo* and the residue purified by flash column chromatography eluting with 2:1 EtOAc:Hex to afford the desired product as a white foam (3.27 g, 85%, mixture of N7/N9 regioisomers). **5**: *R*f = 0.29, 0.36 (silica gel, 97:3 DCM:MeOH); ^{1}H NMR (400 MHz, $CDCl_3$): δ 8.41 (s, 0.5H), 8.39 (d, *J* = 0.8 Hz, 0.5H), 8.38 (s, 1H), 8.32 (d, *J* = 0.8 Hz, 1H), 2.73 (d, *J* = 0.8 Hz, 3H), 2.63 (d, *J* = 0.8 Hz, 1.5H), 1.64 (s, 9H), 1.60 (s, 4.4H), 1.39 – 1.29 (m, 17.8H) ppm; ^{13}C NMR (101 MHz, $CDCl_3$) δ 181.62, 181.46, 151.92, 146.67, 146.31, 145.86, 145.27, 144.91, 143.76, 138.26, 137.65, 137.47, 130.87, 126.75, 121.55, 86.96, 86.66, 81.05, 47.98, 46.73, 28.05, 27.98, 27.96, 27.81, 22.63, 21.66, 21.51, 18.87, 13.67 ppm; HRMS (ESI-TOF) calcd for $C_{20}H_{27}N_4O_4H^+$ $[M+H]^+$ 387.20323, found: 387.20368.

***tert*-butyl-7-(bromomethyl)-4-(3,3,4,4-tetramethyl-2,5-dioxopyrrolidin-1-yl)-1*H*-imidazo[4,5-*c*]pyridine-1-carboxylate** and ***tert*-butyl 7-(bromomethyl)-4-(3,3,4,4-tetramethyl-2,5-dioxopyrrolidin-1-yl)-3*H*-imidazo[4,5-*c*]pyridine-3-carboxylate (11)**[7]

To a flask containing **10** (3.27 g, 8.46 mmol) was added *N*-bromosuccinimide (1.28 g, 7.2 mmol, recrystallized from hot H_2O), benzoyl peroxide (131 mg, 0.54 mmol), and 175 mL of CCl_4 (reagent grade, not anhydrous). The mixture was heated to reflux with stirring. After 3 h additional NBS (301 mg, 1.69 mmol) and $(BzO)_2$ (31 mg, 0.127 mmol) were added. The reaction was carefully monitored by TLC (97:3 DCM:MeOH) afterwards. After 6.5 h total the reaction was virtually complete, with only a trace of starting material

and a trace of undesired dibrominated product. Volatiles were removed *in vacuo* and the residue purified by flash column chromatography eluting with 1:1 Hex:EtOAc to afford the desired product as a white foam (2.78 g, 71%, mixture of N7/N9 regioisomers), as well as the dibrominated product as an off-white solid (357 mg, 7.7%, mixture of N7/N9 regioisomers). **11**: *R*f = 0.36, 0.43 (silica gel, 97:3 DCM:MeOH); ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.61 (s, 1H), 8.50 (s, 1H), 8.49 (s, 0.5H), 8.41 (s, 0.5H), 5.13 (s, 1H), 4.90 (s, 2H), 1.70 (s, 4.5H), 1.62 (s, 9H), 1.42 – 1.30 (m, 18H) ppm; ^{13}C NMR (101 MHz, $CDCl_3$) δ 181.27, 181.08, 151.19, 146.01, 145.47, 144.84, 143.69, 139.70, 132.91, 126.07, 125.29, 121.49, 87.70, 87.03, 48.02, 46.78, 29.46, 27.92, 27.88, 27.85, 23.84, 22.45, 21.51, 21.37, 18.80 ppm; HRMS (DART-TOF) calcd for $C_{20}H_{26}N_4O_4Br$ [M+H]+ 467.11170, found: 467.11272.

***N*-(7-((*tert*-butyldimethylsilyloxy)methyl)-1*H*-imidazo[4,5-*c*]pyridin-4-yl)-2,2,3,3-tetramethylsuccinimide (12)**[7] To a flask containing **11** (1.117 g, 2.406 mmol) and NaOAc (986 mg, 12.03 mmol) was added DMF (17 mL). The mixture was heated to 60-65 °C with stirring. Another 395 mg (4.812 mmol) of NaOAc was added after 4 h, and again after 6 h. After TLC indicated complete consumption of starting material, volatiles were removed *in vacuo* and 7 M NH_3/MeOH (25 mL) was added. The mixture was allowed to stir at rt for 8 h until TLC indicated complete conversion to the free alcohol. Volatiles were removed *in vacuo* and the crude product was loaded onto a flash column. Elution with 9:1 DCM:MeOH afforded 639 mg (88%) of as a white amorphous solid. **Free Alcohol,** *N*-(7-(hydroxymethyl)-1*H*-imidazo[4,5-*c*]pyridin-4-yl)-2,2,3,3-tetramethylsuccinimide: *R*f = 0.16 (silica gel, 92.5:7.5 DCM:MeOH); ^{1}H NMR (CD3OD,

500 MHz) δ 8.37 (s, 1H), 8.31 (s, 1H), 5.02 (s, 2H), 1.38 (s, 12H) ppm; ^{13}C NMR (101 MHz, CD_3OD) δ 183.13, 172.37, 146.58, 141.83, 118.92, 61.65, 21.85, 20.63 ppm; HRMS (ESI-TOF) calcd for $C_{15}H_{19}N_4O_3H$ $[M+H]^+$ 303.14571, found: 303.14639.

To a flask containing the free alcohol (100 mg, 0.331 mmol), TBSCl 69 mg, 0.463 mmol), and DMAP (2 mg, 0.0165 mmol) was added DCM (1.6 mL), DMF (0.4 mL), and DIPEA (86 μL, 0.496 mmol). The mixture was allowed to stir at rt for 20 h, then concentrated *in vacuo* and redissolved in EtOAc (10 mL). The organic layer was washed with water (2 x 5 mL) and brine (5 mL), then dried over Na_2SO_4, filtered, and evaporated. The residue was purified by flash chromatography eluting with 97:3 DCM:MeOH to obtain 117 mg (85%) of **12** as a white amorphous solid. **12**: *R*f = 0.36 (silica gel, 95:5 DCM:MeOH); ^{1}H NMR (500 MHz, CD_3OD) δ 8.38 (s, 1H), 8.35 (s, 1H), 5.17 (s, 2H), 1.38 (s, 12H), 0.96 (s, 9H), 0.15 (s, 6H) ppm; ^{13}C NMR (CD_3OD, 126 MHz) δ 183.26, 146.30, 139.31, 61.04, 26.33, 21.87, 19.22, -5.19 ppm; HRMS (ESI-TOF) calcd for $C_{21}H_{31}N_4O_3SiNa$ $[M+Na]^+$ 439.2141, found: 439.2138.

***N*-(7-(azidomethyl)-1*H*-imidazo[4,5-*c*]pyridin-4-yl)-2,2,3,3-tetramethylsuccinimide (13)**[7] To a flask containing **11** (800 mg, 1,72mmol) and NaN_3 (450 mg, 6.89 mmol) was added DMF (34 mL). The mixture was heated to 45 °C with stirring. After 1.5 h, TLC indicated complete consumption of starting material. Volatiles were removed *in vacuo* and loaded onto a flash column. Elution with 95:5 DCM:MeOH afforded 559 mg (99%) of **13** as a white amorphous solid. ^{1}H NMR (acetone-d_6, 500 MHz) δ 8.37 (s, 1H), 8.34 (s, 1H), 4.92 (s, 2H), 1.35 (s, 12H) ppm; ^{13}C NMR (101 MHz, acetone-d_6) δ 180.75, 144.28,

140.26, 118.92, 48.53, 47.51, 20.78 ppm; HRMS (ESI-TOF) calcd for $C_{15}H_{18}N_7O_2^+$ [M + H]$^+$ 328.1522, found: 328.1536.

General Method for Glycosylation Reactions

To a flask containing the desired heterocycle (**8, 9, 12, 13**) and NaH (1.5 eq) was added MeCN (final concentration of heterocycle = 0.05 M). The mixture was allowed to stir for 0.5 h followed by addition of α-chlorosugar **3** (1.8 eq). The reaction stirred for 2-8 hr, after which the reaction mixture was filtered through a celite pad. Volatiles were removed *in vacuo* and the resulting crude product was purified by flash chromatography. All derivatives can be purified with 1:1 Hex:EtOAc to afford the glycosylated product as a white foam.

14a: Yield – 64%. ^{1}H NMR (400 MHz, acetone-d_6): δ 8.49 (s, 1H), 8.20 (d, J = 5.6 Hz, 1H), 8.02 (d, J = 6.6 Hz, 2H), 7.92 (d, J = 6.4 Hz, 2H), 7.86 (d, J = 5.6 Hz, 1H), 7.38 (d, J = 8.1 Hz, 2H), 7.32 (d, J = 8.4 Hz, 2H), 6.76 (dd, J = 2.8 Hz, 5.8 Hz, 1H), 5.87 (d, J = 6.8 Hz, 1H), 4.71(s, 3H), 3.18-3.12 (m, 1H), 3.06-3.04 (m, 1H), 2.44 (s, 3H), 2.41 (s, 3H), 1.34 (d, J = 15.2 Hz, 12H). ^{13}C NMR (100 MHz, acetone-d_6): δ 180.87, 165.70, 165.52, 144.27, 144.02, 143.83, 141.16, 139.82, 139.28, 138.03, 129.59, 129.52, 129.29, 129.22, 127.10, 127.00, 108.40, 86.032, 82.65, 74.87, 64.03, 47.59, 37.42, 20.78, 20.74 ppm. HRMS (ESI+): calc. 625.26622 found 625.26527.

14b: yield – 74%. *R*f = 0.35 (silica gel, 97:3 DCM:MeOH); ^{1}H NMR (acetone-*d*6, 600 MHz) δ 8.56 (s, 1H), 8.13 (d, *J* = 0.9 Hz, 1H), 8.05 – 7.96 (m, 2H), 7.93 – 7.83 (m, 2H), 7.40 – 7.33 (m, 2H), 7.33 – 7.24 (m, 2H), 6.90 (dd, *J* = 7.8, 5.8 Hz, 1H), 5.86 (dt, *J* = 6.3,

3.1 Hz, 1H), 4.76 (td, *J* = 4.7, 3.2 Hz, 1H), 4.71 – 4.60 (m, 2H), 3.26 (ddd, *J* = 14.4, 7.8, 6.7 Hz, 1H), 3.09 (ddd, *J* = 14.3, 5.8, 3.0 Hz, 1H), 2.81 (s, 3H), 2.43 (s, 3H), 2.39 (s, 3H), 1.33 (d, *J* = 32.6 Hz, 12H) ppm; ^{13}C NMR (acetone-*d*6, 151 MHz) δ 181.76, 166.50, 166.36, 145.18, 144.78, 143.51, 143.49, 140.30, 138.59, 138.34, 130.55, 130.38, 130.13, 130.11, 27.99, 127.91, 119.95, 86.26, 83.39, 75.55, 64.81, 48.39, 38.59, 21.93, 21.72, 21.61, 21.57, 21.44, 15.51 ppm; HRMS (ESI-TOF) calcd for $C_{36}H_{38}N_4O_7Na^+$ $[M + Na]^+$ 661.2638, found: 661.2629.

14c: yield – 76%. *R*f = 0.35 (silica gel, 97:3 DCM:MeOH); ^{1}H NMR (acetone-*d*6, 600 MHz) δ 8.56 (s, 1H), 8.13 (d, *J* = 0.9 Hz, 1H), 8.05 – 7.96 (m, 2H), 7.93 – 7.83 (m, 2H), 7.40 – 7.33 (m, 2H), 7.33 – 7.24 (m, 2H), 6.90 (dd, *J* = 7.8, 5.8 Hz, 1H), 5.86 (dt, *J* = 6.3, 3.1 Hz, 1H), 4.76 (td, *J* = 4.7, 3.2 Hz, 1H), 4.71 – 4.60 (m, 2H), 3.26 (ddd, *J* = 14.4, 7.8, 6.7 Hz, 1H), 3.09 (ddd, *J* = 14.3, 5.8, 3.0 Hz, 1H), 2.81 (s, 3H), 2.43 (s, 3H), 2.39 (s, 3H), 1.33 (d, *J* = 32.6 Hz, 12H) ppm; ^{13}C NMR (acetone-*d*6, 151 MHz) δ 181.76, 166.50, 166.36, 145.18, 144.78, 143.51, 143.49, 140.30, 138.59, 138.34, 130.55, 130.38, 130.13, 130.11, 27.99, 127.91, 119.95, 86.26, 83.39, 75.55, 64.81, 48.39, 38.59, 21.93, 21.72, 21.61, 21.57, 21.44, 15.51 ppm; HRMS (ESI-TOF) calcd for $C_{36}H_{38}N_4O_7Na^+$ $[M + Na]^+$ 661.2638, found: 661.2629.

14d: yield – 74%. ^{1}H NMR (acetone-*d*6, 500 MHz) δ 8.65 (s, 1H), 8.42 (s, 1H), 8.02 (d, *J* = 6.5 Hz, 2H), 7.87 (d, *J* = 5.0 Hz, 2H), 7.37 (d, *J* = 10.0 Hz, 2 H), 7.28 (d, *J* = 5.0 Hz, 2H), 6.85 (dd, *J* = 10, 5.0 Hz, 1H), 5.88 (dt, *J* = 6.0, 3.0 Hz, 1H), 5.03 (quartet, *J* = 14.5, 14.5 Hz, 2H), 4.76 (td, *J* = 4.7, 3.2 Hz, 1H), 4.65 – 4.64 (m, 2H), 3.29 (m, 1H), 3.12 (m, 1H), 2.43 (s, 3H), 2.39 (s, 3H), 1.35 (d, *J* = 27.5 Hz, 12H) ppm; ^{13}C NMR (acetone-*d*6,

125 MHz) δ 166.50, 166.34, 145.21, 144.79, 144.19, 143.75, 143.36, 141.08, 140.00, 139.07, 130.60, 130.57, 130.49, 130.37, 130.14, 130.11, 127.99, 127.92, 118.09, 88.39, 86.65, 85.49, 83.59, 76.09, 75.60, 64.957, 64.80, 50.01, 48.55, 38.42, 30.29, 21.91, 21.61, 21.57, 21.40 ppm. HRMS (ESI-TOF) calcd for $C_{36}H_{38}N_7O_7$ $[M + H]^+$ 680.28327, found: 680.28212.

General Method for Toluoyl Deprotection

To an oven-dried flask containing **14a-d** (0.234 mmol) was added MeOH containing 7 N NH_3 (7 mL). The flask was tightly stoppered and the mixture allowed to stir for 20 h. TLC indicated complete conversion to the free sugar. Volatiles were removed *in vacuo* and the residue purified by flash chromatography eluting with 9:1 DCM:MeOH to afford the product as a white foam (53 mg, 58%).

15a: yield – 58%. ^{1}H NMR (500 MHz, acetone-d_6) δ 8.52 (s, 1H), 8.32 (d, *J* = 5.5 Hz, 1H), 7.89 (d, *J* = 6.0 Hz, 1H), 6.48 (dd, *J* = 1.5 and 6.0, 1H), 4.66 (s, 1H), 4.63 (s, 1H), 4.06 (q, *J* = 3.5 Hz, 1H), 3.79 (s, 2H), 2.77-2.71 (m, 1H), 2.56-2.53 (m, 1H), 1.34 (d, *J* = 21, 12H) ppm. ^{13}C NMR (100 MHz, acetone-d_6): δ 180.95, 144.10, 141.00, 140.14, 139.05, 137.93, 108.33, 88.28, 85.78, 71.20, 61.90, 47.58, 40.55, 20.93, 20.70 ppm. HRMS (ESI+): calc. 389.18249 found 389.18432.

15b: yield – 82%. *R*f = 0.41 (silica gel, 92.5:7.5 DCM:MeOH); ^{1}H NMR (400 MHz, MeOH-*d4*) δ 8.73 (s, 1H), 8.18 (s, 1H), 6.71 (t, *J* = 6.3 Hz, 1H), 4.59 (dt, *J* = 6.1, 4.1 Hz, 1H), 4.07 (q, *J* = 3.7 Hz, 1H), 3.74 (ddd, *J* = 28.9, 12.1, 3.7 Hz, 2H), 2.87 – 2.73 (m, 1H), 2.80 (s, 3H), 2.63 – 2.51 (m, 1H), 1.40 (s, 12H) ppm; ^{13}C NMR (MeOH-*d4*, 101 MHz) δ 183.41, 145.50, 143.29, 141.25, 138.18, 137.73, 121.32, 89.40, 86.97, 71.95, 62.65,

41.90, 30.67, 29.53, 21.89, 21.85, 15.69 ppm; HRMS (ESI-TOF) calcd for $C_{20}H_{26}N_4O_5Na$ $[M+Na]^+$ 425.1801, found: 425.1794.

15c: yield – 86%. *R*f = 0.27 (silica gel, 94:6 DCM:MeOH); ^{1}H NMR (MeOH-*d4*, 500 MHz) δ 8.77 (s, 1H), 8.33 (s, 1H), 6.86 (t, *J* = 6.3 Hz, 1H), 5.17 (dd, *J* = 66.6, 12.8 Hz, 2H), 4.59 (dt, *J* = 6.1, 4.1 Hz, 1H), 4.03 (q, *J* = 3.6 Hz, 1H), 3.72 (ddd, *J* = 33.7, 12.1, 3.6 Hz, 2H), 2.82–2.71 (m, 1H), 2.54 (ddd, *J* = 13.5, 6.2, 4.4 Hz, 1H), 1.38 (brs, 12H), 0.98 – 0.86 (m, 9H), 0.16 (s, 3H), 0.14 (s, 3H) ppm; ^{13}C NMR (MeOH-*d4*, 126 MHz) δ 183.29, 145.89, 142.17, 140.89, 139.47, 138.64, 123.94, 89.36, 87.24, 72.02, 62.70, 61.50, 42.28, 26.32, 21.88, 21.82, 19.06, -5.02, -5.03 ppm; HRMS (ESI-TOF) calcd for $C_{26}H_{40}N_4O_6SiNa$ $[M+Na]^+$ 555.2592, found: 555.2603.

19: yield – 73%.^{1}H NMR (500 MHz, acetone-d_6) δ 8.52 (s, 1H), 8.32 (d, *J* = 5.5 Hz, 1H), 7.89 (d, *J* = 6.0 Hz, 1H), 6.48 (dd, *J* = 1.5 and 6.0, 1H), 4.66 (s, 1H), 4.63 (s, 1H), 4.06 (q, *J* = 3.5 Hz, 1H), 3.79 (s, 2H), 2.77-2.71 (m, 1H), 2.56-2.53 (m, 1H), 1.34 (d, *J* = 21, 12H) ppm. ^{13}C NMR (100 MHz, acetone-d_6): δ 180.95, 144.10, 141.00, 140.14, 139.05, 137.93, 108.33, 88.28, 85.78, 71.20, 61.90, 47.58, 40.55, 20.93, 20.70 ppm. HRMS (ESI-TOF): calc. 389.18249 found 389.18432.

General Method for Dimethoxytrityl-Protection

To a flask containing the deprotected sugar (**15a-c, 20**, 0.45 mmol) and DMTr-Cl (1.2 eq, 0.53 mmol) was added pyridine (9 mL). The solution was allowed to stir for 6 hr. TLC (95:5 DCM:MeOH) indicated complete consumption of starting material. Staining with 10% Sulfuric Acid appeared orange at first and charred to black with heat. The reaction

solution was concentrated *in vacuo* and applied to silica gel. Column Chromatography with 98:2:1 DCM:MeOH:TEA afforded the protected nucleoside.

16a: yield – 71%. ^{1}H NMR (500 MHz, acetone-*d6*) δ 8.38 (s, 1H), 8.16 (d, *J* = 5.5 Hz, 2H), 7.45 (d, *J* = 5.5 Hz, 2H), 7.33 – 7.20 (m, 7H), 6.84 – 6.82 (m, 4H), 6.72 (d, *J* = 6.0 Hz, 1H), 6.50 (t, *J* = 6.5 Hz, 1H), 4.71 (q, *J* = 3.5 Hz, 1), 4.21 (q, *J* = 5 Hz, 1H), 3.76 (s, 6H), 3.38 (qd, *J* = 10, 5.5 Hz, 2H), 2.80 (m, 1H), 2.60 (m, 1H), 1.34 (d, *J* = 21 Hz, 12H). ^{13}C NMR (acetone-*d6*, 125 MHz) δ 180.88, 158.74, 145.05, 143.86, 140.98, 139.96, 139.16, 138.11, 135.84, 135.69, 130.11, 128.14, 127.76, 126.74, 113.05, 108.70, 86.66, 86.29, 85.80, 71.20, 63.92, 54.64, 47.58, 45.57, 40.15, 38.57, 20.75 ppm; HRMS (MALDI) calcd for $C_{40}H_{43}N_4O_7$ $[M+H]^+$ 691.3126, found: 692.3135.

16b: yield – 87%. *R*f = 0.34 (silica gel, 95:5 DCM:MeOH); ^{1}H NMR (400 MHz, $CDCl_3$/pyridine-*d*5) δ 8.18 (d, *J* = 0.7 Hz, 1H), 8.16 (s, 1H), 7.38 (dd, *J* = 8.2, 1.3 Hz, 2H), 7.31 – 7.11 (m, 7H), 6.79 (d, *J* = 8.9 Hz, 4H), 6.52 (t, *J* = 6.3 Hz, 1H), 4.55 (dd, *J* = 9.0, 4.9 Hz, 1H), 4.16 (q, *J* = 4.5 Hz, 1H), 3.77 (s, 6H), 3.33 (ddd, *J* = 24.2, 10.2, 4.7 Hz, 2H), 2.66 (s, 3H), 2.52 (t, *J* = 5.7 Hz, 2H), 1.37 (d, *J* = 4.6 Hz, 13H) ppm; ^{13}C NMR (101 MHz, $CDCl_3$/pyridine-*d*5) δ 181.72, 181.69, 158.69, 146.05, 144.55, 143.13, 143.08, 142.29, 142.25, 139.55, 137.39, 137.29, 135.68, 135.60, 130.13, 130.11, 130.02, 129.98, 129.14, 128.33, 128.15, 128.05, 127.07, 125.41, 118.50, 113.35, 86.74, 86.30, 85.21, 85.16, 71.63, 63.77, 55.34, 55.28, 47.98, 47.94, 46.24, 41.64, 21.91, 21.86, 21.80, 21.69, 21.54, 21.44, 21.37, 15.87, 0.12 ppm; HRMS (ESI-TOF) calcd for $C_{41}H_{44}N_4O_7Na$ $[M+Na]^+$ 727.3073, found: 727.3094.

16c: yield – 84%. R_f = 0.34 (silica gel, 95:5 DCM:MeOH); ^{1}H NMR (500 MHz, $CDCl_3$/pyridine-d_5) δ 8.32 (s, 1H), 8.22 (s, 1H), 7.45 – 7.33 (m, 2H), 7.33 – 7.11 (m, 7H), 6.87 – 6.69 (m, 5H), 5.02 (dd, *J* = 98.3, 12.8 Hz, 3H), 4.59 (q, *J* = 5.2 Hz, 1H), 4.16 (q, *J* = 4.6 Hz, 1H), 3.76 (s, 6H), 3.34 (qd, *J* = 10.2, 4.6 Hz, 2H), 2.53 (dd, *J* = 8.8, 3.9 Hz, 3H), 1.38 (s, 6H), 1.36 (s, 6H), 0.90 (s, 9H), 0.11 (s 3H), 0.9 (s, 3H) ppm; ^{13}C NMR ($CDCl_3$/pyridine-d_5, 125 MHz) δ 181.4, 158.5, 144.5, 142.5, 141.7, 139.3, 138.9, 137.6, 135.7, 135.5, 130.01, 129.98, 128.1, 127.9, 126.9, 121.3, 113.2, 86.5, 85.9, 85.2, 71.2, 63.7, 60.8, 55.2, 47.9, 46.2, 41.6, 25.8, 21.7, 21.2, 18.2, -5.1, -5.2 ppm; HRMS (ESI-TOF) calcd for $C_{41}H_{44}N_4O_7Na$ $[M+Na]^+$ 857.3922, found: 857.3903.

21: yield – 91%. ^{1}H NMR (acetone-*d6*, 500 MHz) δ 8.36 (s, 1H), 8.24 (s, 1H), 7.85-6.73 (m, 22H), 4.95 (qd, *J* = 19.5, 5 Hz, 2H), 4.45 (m, 2H), 4.29 (t, *J* = 10 Hz, 1H), 4.23 (t, *J* = 7.5 Hz, 1H), 4.10 (bs, 1H), 3.75 (s, 3H), 3.71 (s, 3H), 3.11-2.95 (m, 2H), 2.41 (m, 1H), 2.32 (m, 1H), 1.35 (d, *J* = 21.5 Hz, 12H). ^{13}C NMR (acetone-*d6*, 126 MHz) δ 158.90, 158.84, 158.65, 145.56, 144.88, 144.15, 144.10, 143.10, 141.22, 141.16, 138.81, 137.94, 136.28, 136.17, 135.71, 135.70, 130.35, 130.20, 129.95, 128.27, 128.01, 127.87, 127.69, 127.62, 127.60, 127.02, 126.84, 126.65, 125.35, 125.27, 119.87, 119.84, 113.23, 112.99, 112.97, 87.13, 86.21, 86.05, 74.83, 66.53, 63.98, 54.62, 54.60, 47.58, 47.18, 39.78, 20.99, 20.82, 20.71, 20.57 ppm. HRMS (ESI-TOF) calcd for $C_{56}H_{55}N_5O_9Na$ $[M+Na]^+$: 964.3897, found 964.3896

General Method for Phosphitylation

To a flask containing the 5'-*O*-DMT-protected nucleoside (**16a-c or 21**, 0.403 mmol) dissolved in MeCN (8 mL) at 0 ^{o}C was added tetrazole (0.90 mL of 0.45 M solution in MeCN, 0.403 mmol) followed by 2-cyanoethyl-*N*,*N*,*N*,*N*-tetraisopropylphosphane (256 µL, 0.807 mmol) dropwise. Conversely, *N,N*-Diisopropylamino-2-cyanoethylphosphonamidic chloride (0.807 mmol) can be used without tetrazole as an activator. The mixture was allowed to warm up and stir at rt for 6 h, after which TLC indicated that the reaction was complete. Methanol (1 mL) was added and stirring continued for 10 min. The reaction was concentrated *in vacuo* and applied to silica gel column chromatography (gradient of 1:1 Hex:EtOAc to 100% EtOAc). Precipitation by slow addition of the crude mixture dissolved in minimal DCM to vigorously stirring hexanes at 0 °C afforded the desired phosphoramidite (**17a-c, 22**) as a white solid (1:1 mixture of diasteriomers).

17a: yield – 73%. ^{31}P NMR (203 MHz, $CDCl_3$) = 147.36, 146.95 ppm; HRMS (MALDI-TOF) calcd for $C_{49}H_{59}N_6O_8PK$ $[M+K]^+$ calc. 930.1002, found: 930.1752.

17b: yield – 85%. *R*f = 0.41, 0.44 (silica gel, 97:3 DCM:MeOH); ^{1}H NMR (400 MHz, $CDCl_3$/pyridine-*d*5) δ 8.21 (s, 1.5H), 8.18 (s, 0.5H), 7.44 – 7.33 (m, 2H), 7.33 – 7.15 (m, 7H), 6.86 – 6.74 (m, 4H), 6.53 (dd, *J* = 7.4, 5.7 Hz, 1H), 4.72 – 4.55 (m, 1H), 4.34 – 4.20 (m, 1H), 3.86 – 3.56 (m, 4H), 3.78 (s, 3H), 3.77 (s, 3H), 3.39 – 3.22 (m, 2H), 2.69 (dd, *J* = 4.6, 0.6 Hz, 3H), 2.67 – 2.50 (m, 3H), 2.44 (t, *J* = 6.3 Hz, 1H), 1.38 (s, 6H), 1.37 (s, 6H), 1.23 – 1.14 (m, 9H), 1.11 (d, *J* = 6.8 Hz, 3H) ppm; ^{31}P NMR (162 MHz, $CDCl_3$/pyridine-*d*5) δ 149.44, 149.02 ppm; HRMS (ESI-TOF) calcd for $C_{50}H_{61}N_6O_8PNa$ $[M+Na]^+$ 927.4186, found: 927.4208.

17c: *R*f = 0.41, 0.44 (silica gel, 97:3 DCM:MeOH); ^{1}H NMR (500 MHz, $CDCl_3$/pyridine-*d5*) = 8.32 (s, 1H), 8.22 (s, 1H), 7.45 – 7.33 (m, 2H), 7.33 – 7.11 (m, 7H), 6.87 – 6.69 (m, 5H), 5.02 (dd, J = 98.3, 12.8 Hz, 3H), 4.59 (q, *J* = 5.2 Hz, 1H), 4.16 (q, *J* = 4.6 Hz, 1H), 3.76 (d, *J* = 0.8 Hz, 6H), 3.34 (qd, *J* = 10.2, 4.6 Hz, 2H), 2.53 (dd, *J* = 8.8, 3.9 Hz, 3H), 1.37 (d, *J* = 7.0 Hz, 12H), 0.90 (s, 9H), 0.11 (s, 3H), 0.09 (s, 3H) ppm; ^{31}P NMR (203 MHz, $CDCl_3$/pyridine-*d5*) = 149.51, 149.42 ppm; HRMS (ESI-TOF) calcd for $C_{56}H_{76}N_6O_9PSi$ $[M+H]^+$ calc. 1035.5181, found: 1035.5179.

22: yield – 73%. ^{31}P NMR (acetone-*d6*) δ 148.38, 148.31. HRMS (ESI-TOF): calcd for $C_{65}H_{72}N_7O_{10}PNa$ $[M+Na]^+$ 1164.4976, found 1164.4944.

1-[2-deoxy-3,5-di-*O*-(*p*-toluoyl)-β-D-*erythro*-pentofuranosyl]-4-(3,3,4,4-tetramethyl-2,5-dioxopyrrolidin-1-yl)-7-aminomethyl-1*H*-imidazo[4,5-*c*]pyridine (18). To a flask containing **17** (725 mg, 1.07 mmol) was added Pd/C (110 mg, 10% wt Pd on C) and 21 mL of EtOAc. Hydrogen gas was bubbled through the solution for 20 minutes and then allowed to stir under an H_2 atmosphere for 20 hr. TLC (95:5 DCM:MeOH) indicated complete consumption of starting material. Staining with 0.2% Ninhydrin in EtOH verified the presence of a free amine. The suspension was filtered through celite and the filtrate concentrated. Column chromatography eluting with 2.5% MeOH, 1% TEA, and 97.5% DCM afforded 458 mg of **30** in 78% yield. ^{1}H NMR (acetone-*d*6, 500 MHz) δ 8.58 (s, 1H), 8.27 (s, 1H), 8.00 (d, *J* = 8.5 Hz, 2H), 7.93 (d, *J* = 8.5 Hz, 2H), 7.39 (d, *J* = 8.0 Hz, 2H), 7.32 (d, *J* = 8.5 Hz, 2H), 7.15 (dd, *J* = 14, 6 Hz, 1H), 5.84 (dt, *J* = 6.0, 3.0 Hz, 1H), 5.04 (d, *J* = 15.5 Hz, 1H), 4.80 (d, *J* = 15.5 Hz, 1H), 4.71-4.65 (m, 3H), 3.12 – 3.04 (m, 2H), 2.44 (s, 3H), 2.39 (s, 3H), 1.36 (d, *J* = 29 Hz, 12H) ppm. ^{13}C NMR

(acetone-*d*6, 125 MHz) δ 165.69, 165.37, 144.36, 143.94, 142.90, 141.87, 139.45, 138.78, 137.88, 129.62, 129.52, 129.26, 129.24, 127.16, 127.11, 122.14, 86.18, 82.45, 75.11, 64.16, 50.48, 47.55, 38.30, 21.02, 20.85, 20.73, 20.69, 20.58. HRMS (ESI-TOF) calcd for $C_{36}H_{40}N_5O_7$ $[M+H]^+$: 654.29277, found: 654.29418.

1-[2-deoxy-β-D-*erythro*-pentofuranosyl]-4-(3,3,4,4-tetramethyl-2,5-dioxopyrrolidin-1-yl)-7-(fluorenylmethoxycarbonyl)aminomethyl-1*H*-imidazo[4,5-*c*]pyridine (20): To a flask containing **19** (213 mg, 0.511 mmol) in dioxane (776 uL) and 10% $NaHCO_3$ (1.6 mL) was added a solution of Fmoc-Cl (170 mg, 0.660 mmol) in dioxane (1.53 mL). The solution stirred for 6 hr. TLC (92:8 DCM:MeOH) indicated complete consumption of starting material. The reaction mixture was diluted with EtOAc (20 mL), washed with saturated $NaHCO_3$ (3 x 15 mL), and dried over Na_2SO_4. The organic phase was filtered, concentrated *in vacuo*, and purified with column chromatography (92:8 DCM:MeOH) to afford **32** (285 mg, 85%) as a white foam. ^{1}H NMR (acetone-*d6*, 500 MHz) δ 8.63 (s, 1H), 8.34 (s, 1H), 7.85 – 7.27 (m, 8H), 6.69 (t, *J* = 5 Hz, 1H), 4.95 – 4.79 (m, 2H), 4.65 (t, *J* = 3.5 Hz, 1H), 4.39 (d, *J* = 7 Hz, 2H), 4.27 (t, *J* = 4.5 Hz, 1H), 4.08 (q, *J* = 3.5 Hz, 3.5 Hz, 1H), 3.77 (m, 2H), 2.88 (m, 1H), 2.66 (m, 1H), 1.37 (d, *J* = 26.5 Hz, 12H) ppm. ^{13}C NMR (acetone-*d*6, 125 MHz) δ 180.96, 144.22, 144.17, 143.68, 142.28, 141.93, 138.78, 138.60, 137.88, 127.60, 127.05, 125.26, 119.96, 119.88, 119.86, 88.30, 85.73, 70.74, 66.44, 61.72, 47.55, 47.15, 45.80, 40.85, 39.51, 20.92, 20.65, 7.92 ppm. HRMS (ESI-TOF) calcd for $C_{35}H_{37}N_5O_7Na$ $[M+Na]^+$: 662.6907, found 662.1632

DNA Synthesis.

The core sequence chosen for this study d(CCGG AAAA CGCC)/d(GGCG TTTT

CCGG) is the same as that used in related studies.[8]

The native and modified 12-mers were prepared by solid-phase DNA synthesis and deprotected using standard protocols except those containing hmc_3A (**17c**). The analogue nucleotides could be incorporated into DNA strands with essentially the same coupling efficiency as the common nucleotides except for hmc3A, which coupled in 70-80% yield using the standard wait time of *2 min* during the coupling cycle. Increasing the wait step of the coupling cycle to *1 hour* for each introduction of **17c** resulted in coupling efficiencies of 85–90%. All newly-synthesized oligos were deprotected and cleaved from the CPG beads by 12–14 h treatment with concentrated ammonia at 55 °C. The solutions were then decanted and evaporated by speedvac, then redissolved in 4-6 mL H_2O for HPLC purification.

Initial purification of native and modified oligonucleotides sequences 1–8 was accomplished by HPLC (Oligo R3 reverse-phase C18 column, trityl on), starting with 100% A using a linear gradient from 0 to 50% B over 28 min [A: 1 M TEAA (pH 7) in 5% acetonitrile; B: 1 M TEAA (pH 7) in 70% acetonitrile]. The DMT-protected 12-mers had retention times of about 17.5 minutes, except for Sequence 7, which had a retention time of 24 min due to the presence of a TBS group remaining on the hydroxymethyl group of each residue of **17c** (four total). The collected oligonucleotides were then reduced in volume and detritylated and/or desilylated.

All sequences except those containing hmc3A were detritylated in 80% acetic acid for 30 min at 0 °C. The resulting oligonucleotides were then desalted (Sephadex G-

10) and lyophilized. Sequences containing hmc3A were not detritylated with AcOH because the TBS groups were not fully cleaved after 1 h treatment at rt. Instead, after DMT-on HPLC purification the sequences were lyophilized and dissolved in 100 µL anhydrous DMSO. Then 125 µL of NEt_3•3HF were added and the mixture heated at 65 °C for 2.5 h (DMT-off conditions). Desalting of hmc3A sequences was accomplished by precipitation from butanol: after cooling the desilyation mixture briefly in a freezer, 1 mL of 1-butanol was added, followed by 25 µL of 3 M NaOAc. After vortexing for 30s the mixture was cooled at -70°C for 30 minutes. This was followed by centrifugation for 10 min at 12,500 rpm. The butanol was decanted using a sterile pipette. The residue was rinsed with 0.75 mL ethanol twice, then dried under high vacuum in a speed-vac to remove traces of butanol. Sequences containing 4 hmc_3A were purified again by HPLC (Oligo R3 as above) before use; sequences with 1 hmc_3A wer used without further purification.

Nucleic Acid Analysis

MALDI-TOF mass spectrometry was used to obtain masses of fully deprotected oligos. All MALDI-TOF spectra were collected on a Waters MALDI micro MX spectrometer in positive-ion, linear mode. The samples were prepared with 1 uL of 200 uM stock solution of each oligonucleotide, 1 uL of ammonium citrate buffer (10 mM, pH = 9.4), 1 uL of 200 uM DNA standard (5'-GCT GAA TAC ATA AGA CG-3'), and 6 uL of saturated 3-hydroxypiccolinic acid (Sigma Aldrich). The samples were desalted with an ion exchange resin (Dowex 50WX8-400, NH_4^+ form) and spotted onto a MALDI plate

where they were dried at room temperature. The resulting spectra were calibrated relative to the +1 and +2 ions of the internal DNA standard. The results are presented below:

Entry	Sequence	m/z calc	m/z found
1	d(CCGG AAAA CGCC)	3623.7	3623.5
2	d(CCGG AA$\mathbf{A^{H}}$A CGCC)	3622.7	3622.9
3	d(CCGG AA$\mathbf{A^{CH3}}$A CGCC)	3636.7	3636.8
4	d(CCGG AA$\mathbf{A^{CH2OH}}$A CGCC)	3652.7	3652.7
5	d(CCGG AA$\mathbf{A^{CH2NH3+}}$A CGCC)	3651.7	3652.6
6	d(CCGG $\mathbf{A^{H}A^{H}A^{H}A^{H}}$ CGCC)	3619.7	3619.9
7	d(CCGG $\mathbf{A^{CH3}A^{CH3}A^{CH3}A^{CH3}}$ CGCC)	3674.8	3674.8
8	d(CCGG $\mathbf{A^{CH2OH}A^{CH2OH}A^{CH2OH}A^{CH2OH}}$ CGCC)	3739.7	3739.8
9	d(CCGG $\mathbf{A^{CH2NH3+}A^{CH2NH3+}A^{CH2NH3+}A^{CH2NH3+}}$ CGCC)	3737.6	3738.9
10	d(CCGG $\mathbf{A^{CH2NH3+}}$AA$\mathbf{^{CH2NH3+}}$A CGCC)	3677.8	3679.6
11	d(GGCG TTTT CCGG)	3667.6	3667.0

Thermal Denaturation Studies

Thermal denaturation studies were performed in solutions of 20 mM NaH_2PO_4 (pH 7.0) and 1 M NaCl with the concentrations of duplex at 5 μM). Absorbance and temperature values were measured with an AVIV 14DS UV-visible spectrophotometer equipped with digital temperature control. The temperature of the cell compartment was increased in 1.0 °C/min steps (from 4 to 95 °C), and when equilibrium was reached,

temperature and absorbance data were collected every 1 °C. The raw data was exported to the software application Microcal Origin (OriginLab Corporation, Northampton, MA) for determination of the melting temperature (T_m) and free energy (ΔG) of each modified duplex. First, the T_m plot (absorbance vs temperature) was converted to an alpha (α) plot (fraction of molecules paired vs temperature) using the equation:

$$\alpha = \frac{A - A_s}{A_d - A_s} \quad (1)$$

where A is the absorbance at a given temperature, and A_s and A_d are the absorbance values for the single- and double-stranded states, respectively. The parameter α represents the fraction of molecules paired; the T_m is extrapolated when $\alpha = 0.5$. In order to obtain sufficient approximations for A_s and A_d, the upper and lower baselines of the melting curve were fitted using linear least-squares fits. The thermodynamic parameters ΔH, ΔS, and ΔG were determined using a van't Hoff plot (ln K vs $1/T$). The value of K was calculated at each temperature from the α plot using the following equation:

$$K = \frac{\alpha}{2(1-\alpha)^2 C_t} \quad (2)$$

where C_t is the total strand concentration. The thermodynamic parameters can be derived directly from the linear least-squares fit in which the slope is $-\Delta H^o/R$ and the intercept is $\Delta S^o/R$. The free energy change (ΔG^o) at any temperature can be determined by using the following relationship:

$$\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ \quad (3)$$

CD Experiments

CD spectra were collected on an AVIV Circular Dichroism Spectrometer (Model 420). Samples were prepared in a buffer containing 20 mM Phosphate Buffer, pH 7 and 1 M NaCl at a concentration of 3 uM duplex and annealed. Samples were scanned over a range 350 to 220 nm at a rate of with time constant and bandwidth parameters set at 1 s and 1 nm, respectively. Data were collected at 1 nm increments for each scan and an average of 3 scans. The collected data was subtracted from a buffer blank to generate each data set.

References

1) Reese, Colin B. *Organic & Biomolecular Chemistry*, **2005***, 3,* 3851.

2) Reich, M. F.; Fabio, P. F.; Lee, V. J.; Kuck, N. A.; Testa, R. T. *J. Med. Chem.* **1989**, *32*, 2474–2485.

3) Campbell, J. B.; Greene, J. M.; Lavagnino, E. R.; Gardner, D. N.; Pike, A. J.; Snoddy, J.; Taylor, E. C. *J. Heterocyclic Chem.* **1986**, *23,* 669–672.

4) Šilhár, P.; Pohl, R.; Votruba, I.; Hocek, M. *Org. Lett.* **2004**, *6* (19), 3225–3228

5) Gundersen, L-L. *Targets in Heterocyclic Systems* **2008**, *12*, 85-119 and references cited therein.

6) Crey-Desbiolles, C.; Kotera, M. *Bioorg. Med. Chem.* **2006**, *14*, 1935-1941.

7) Lan, T.; McLaughlin, L. W. *J. Am. Chem. Soc.* **2000**, *122*, 6512-6513

8) Salandria, K. J., Arico, J. W., Calhoun, A. K., McLaughlin, L. W. *J. Am. Chem. Soc.* **2011**, *133*, 1766-1768. Reprinted with permission from *J. Am. Chem. Soc.*, 2011, 133 (6), pp 1766–1768. Copyright 2011 American Chemical Society.

JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

COMMUNICATION

pubs.acs.org/JACS

Stability of DNA Containing a Structural Water Mimic in an A-T Rich Sequence

Kerry J. Salandria,[†] Joseph W. Arico,[†] Amy K. Calhoun, and Larry W. McLaughlin*

Department of Chemistry, Merkert Chemistry Center, Boston College, 2609 Beacon Street, Chestnut Hill, Massachusetts 02467, United States

Supporting Information

ABSTRACT: We describe here the synthesis and properties of A-T rich DNA containing covalently bound water mimics located in the DNA minor groove.

Hydration in the minor groove of B-form DNA was initially observed in the early crystal structures described by Dickerson and co-workers.[1,2] More recent studies[3,4] have both confirmed the presence of the minor-groove spine of hydration and mapped the primary and secondary shells and hexagonal assemblies of water molecules. The primary water layer is hydrogen-bonded to the adenine N3 nitrogens (and/or thymine O2 oxygens) and then reaches across the minor groove to a second N3 nitrogen or O2 carbonyl in the opposite strand, one base pair displaced. The second water layer bridges the oxygens of the primary hydration layer.

Perhaps more challenging to understand is the contribution of such hydration patterns to duplex conformation and stability. Simple deletions of the adenine N3 nitrogens[5] and/or thymine O2 carbonyls[6] selectively eliminate hydrogen-bonding sites in the minor groove. Studies with such analogue nucleosides present in DNA have reported destabilization of the duplex in spite of the presence of the Watson−Crick functional groups that should permit essentially normal complementary base pairing. CD studies[7] have indicated that in the absence of selected N3 nitrogens and O2 carbonyls, duplex DNA readily undergoes a conformational change to a structure that is very A-like in nature. Also notable are studies showing that the presence of an uncompensated amino[8] or carbonyl[9] group in the minor groove is dramatically destabilizing.

We have designed an analogue nucleoside (Figure 1b) that attempts to mimic the N3-hydrated state of adenine (Figure 1a). Here we report on the synthesis of the analogue and related compounds, their incorporation into DNA, and the stability of duplexes containing one or more of the "covalently hydrated" adenine residues.

Our synthetic strategy was designed to prepare the 3-deaza and 3-deaza-3-methyl derivatives of dA as controls that would help us better understand the effects of the loss of the N3 nitrogen as well as the effects resulting from the introduction of steric bulk into the minor groove by using the methyl-substituted analogue. The final synthetic target was the covalently hydrated analogue, the 3-deaza-3-hydroxymethyl derivative. The syntheses of the 3-deaza and 3-deaza-3-methyl heterocycles are known, and that of the 3-deaza-3-hydroxymethyl derivative was straightforward. The challenge in all of the syntheses involved the glycosylation reactions. As the steric bulk increases (even from hydrogen to methyl) at the 3-position of the corresponding 6-chloropurine (the most common glycosylation target), the ratio of N7- to N9-glycosylated product becomes extremely unfavorable (N7 > N9). We note that the glycosylation of 6-chloro-3-deaza-3-methylpurine was reported[10] to yield the desired N9-glycosylated product in only 20% yield. We have largely solved[11] the difficulties related to effective glycosylation of 3-substituted purines with the introduction of the tetramethylsuccinimide (M_4SI) protecting group at the 6-position of the purine (Scheme 1). This group effectively blocks access to the N7 nitrogen, causing glycosylation to occur predominantly at the desired N9 nitrogen. Notably, with this protecting group the 3-deaza-3-(*O*-tributylsilyl)hydroxymethylpurine derivative could be glycosylated at the N9 nitrogen in 74% isolated yield (Scheme 1). The succinimide is easily hydrolyzed after the glycosylation (or after DNA synthesis) to unmask the 6-amino group, thus avoiding the harsh conditions necessary to convert the 6-chloro derivative to the corresponding amine. After elaboration of the analogue nucleosides into the fully protected phosphoramidites, they could be incorporated into DNA with yields that were sufficient to prepare the desired sequences.

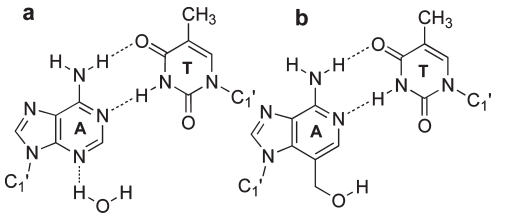
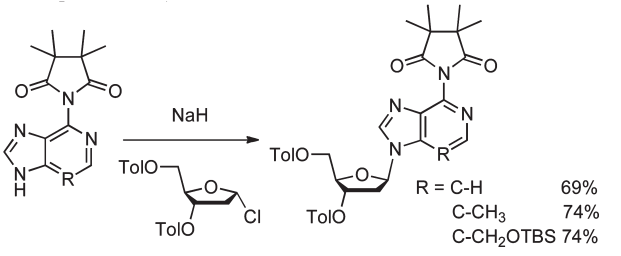


Figure 1. (a) Water molecule hydrogen-bonded to the N3 of dA as part of the minor-groove spine of hydration. (b) Covalently bound water mimic.

Scheme 1. Glycosylation of 3-Substituted Purines Using the Tetramethylsuccinimide (M_4SI) Protecting Group



After ammonia deprotection and purification of the oligonucleotides, the TBS group was removed by dissolving the 12-mers

Received: November 18, 2010
Published: January 18, 2011

ACS Publications © 2011 American Chemical Society

dx.doi.org/10.1021/ja1103684 | *J. Am. Chem. Soc.* 2011, 133, 1766–1768

Table 1. Thermodynamic Stabilities of Selected DNA Duplexes

duplex	sequence at 5 μM duplex	T_M (°C)	ΔG (kcal/mol)
1	d(CCGG AAAA CGCC)	63.4	-19.1 ± 0.6
2	d(CCGG AAA^{H}A CGCC)	61.0	-16.4 ± 0.4
3	d(CCGG AAA^{CH_3}A CGCC)	58.3	-14.8 ± 0.3
4	d(CCGG AAA^{CH_2OH}A CGCC)	60.9	-16.4 ± 2.3
5	d(CCGG $(A^{H})_4$ CGCC)	51.1	-13.9 ± 0.2
6	d(CCGG $(A^{CH_3})_4$ CGCC)	44.4	-10.8 ± 0.4
7	d(CCGG $(A^{CH_2OH})_4$ CGCC)	54.7	-15.2 ± 0.7

in dimethyl sulfoxide followed by treatment with fluoride ion. Confirmation of the complete deprotection of the oligonucleotides was obtained by mass spectral analysis.

We prepared three types of oligonucleotides: (i) those containing 3-deazaadenine, (ii) those containing 3-deaza-3-methyladenine, and (iii) those containing 3-deaza-3-hydroxymethyladenine. The design parameters were as follows: The first modified oligos did not contain the N3 nitrogen, a site that is used in the formation of the spine of hydration but otherwise introduces no unfavorable steric effects. In the methyl-containing sequences, the N3 nitrogen was also removed, and additional unfavorable steric effects were introduced into the minor groove through the presence of one or more methyl groups. In the final sequences, hydroxyls were introduced onto the unfunctionalized methyls to generate a covalently bound water mimic capable of interacting by hydrogen bonding in the minor groove.

Thermodynamic parameters were obtained for the formation of all duplexes (Table 1). Even the simple elimination of a single N3 nitrogen from a centrally located dA residue (duplex **2**) resulted in a 2.4 °C decrease in the melting temperature (T_M), consistent with previous reports.[11] In that same study, the introduction of three dA residues lacking the N3 nitrogen resulted in a 12 °C change in T_M, which also compares favorably with the 12.3 °C reduction in T_M observed here with four deazaadenines present (duplex **5**). We then examined the methyl-substituted sequences (duplexes **3** and **6**). The 3-deaza-3-methyl analogues both lack the N3 nitrogen but also add steric bulk within the minor groove. Introduction of a single methyl group (duplex **3**) reduced the T_M value by 2.7 °C relative to a sequence containing a single 3-deaza analogue (duplex **2**) and by 5.1 °C relative to the unmodified standard. The presence of four analogues containing methyls (duplex **6**) reduced the T_M by a dramatic 19.0 °C, attesting to the duplex instability resulting from increased steric effects in the minor groove. The hydroxymethyl group is sterically even larger than a methyl group and could have an even more dramatic destabilizing effect should the hydroxy groups be unable to take part in advantageous interactions in the minor groove. Introduction of one $-CH_2OH$ group did result in helix destabilization relative to the control sequence (compare the T_M for duplex **4** with that for duplex **1**). However, relative to the single-methyl sequence (duplex **3**), the T_M for the hydroxymethyl sequence was some 2.6 °C higher. This observation is further supported by the sequence with four $-CH_2OH$ groups (duplex **7**), which exhibited a T_M 10.3 °C higher than that for the sequence containing four methyl groups. No significant additional cooperativity appeared to be present for any analogue duplex.

The simplest interpretation of these results is that the absence of hydrogen-bonding acceptors in the minor groove destabilizes the duplex as a result of disruption of the ordered spine of hydration. The presence of four methyl groups also disrupts the spine of hydration but additionally introduces further steric effects and additional destabilization of the duplex. The sequence containing four hydroxymethyl substituents in principle has the greatest steric effects in the minor groove, but the hydroxyl groups appear to provide complementary stabilizing interactions through a modified water structure guided by favorable interactions from the $-CH_2OH$ groups. A possible interaction mimicking the corresponding water molecule is illustrated in Figure 2.

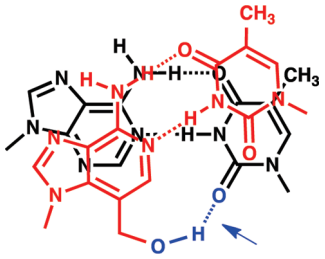


Figure 2. Structural water mimic interacting with the carbonyl of the cross-strand dT one base pair removed.

CD spectra of the various modified helices (Figure 3) also offer some insight into the effects of various analogues. All of the spectra suggest the presence of B-form helices. The CD spectra for the native sequence **1** and sequence **5** containing four dc^{3}A residues are the most alike of the four spectra, suggesting similar chromophoric stacking (dA vs dc^{3}A) within the helix. The interactions of the purine heterocycles are dramatically different for duplex **6**, in which four methyl groups have been added to the central core of base pairs. This spectrum suggests a quite different stacking of the dmc^{3}A chromophores, leading to dramatic shifts in the positive and negative displacements in the observed CD spectrum. The final spectrum, for duplex **7**, is intermediate in character relative to those of duplexes **5** and **6**. It seems likely that the minor-groove interactions present with the structural water mimic drive the conformation back toward a more nativelike B-form, but much like the temperature data, the final native B-like structure is not fully attained.

The data presented here suggest that simple methyl groups located at the C3 position of dA residues are sufficient for dramatic structural perturbation but that the introduction of a $-OH$ group as a structural water mimic assists in partial stabilization of the analogue structure, presumably through interactions within the minor groove.

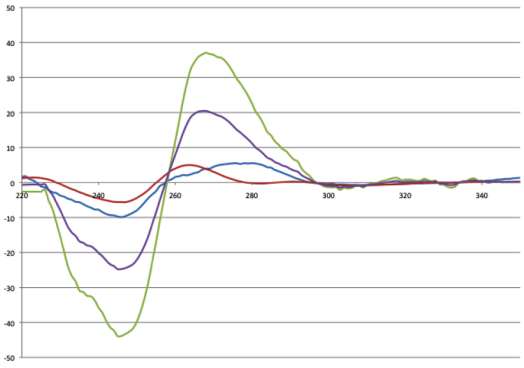


Figure 3. CD spectra (ellipticity vs wavelength at 3 μM duplex concentration) of duplexes **1** (blue), **5** (red), **6** (green), and **7** (purple) obtained in 20 mM sodium phosphate (pH 7) and 1 M NaCl.

dx.doi.org/10.1021/ja1103684 |*J. Am. Chem. Soc.* 2011, 133, 1766–1768

■ ASSOCIATED CONTENT

Supporting Information. Synthetic schemes, procedures, and thermal analyses. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author
mclaughl@bc.edu

Author Contributions
[†]These authors contributed equally.

■ ACKNOWLEDGMENT

This work was supported by the NSF (MCB 0958515). A.K.C. was supported by the Donald T. Moynihan Fund.

■ REFERENCES

(1) Drew, H. R.; Dickerson, R. E. *J. Mol. Biol.* **1981**, *151*, 535–556.
(2) Kopka, M. L.; Fratini, A. V.; Drew, H. R.; Dickerson, R. E. *J. Mol. Biol.* **1982**, *163*, 129–146.
(3) Shui, X. Q.; McFail-Isom, L.; Hu, G. G.; Williams, L. D. *Biochemistry* **1998**, *37*, 8341–8355.
(4) Tereshko, V.; Minasov, G.; Egli, M. *J. Am. Chem. Soc.* **1999**, *121*, 470–471.
(5) Seela, F.; Rosemeyer, H.; Fischer, S. *Helv. Chim. Acta* **1990**, *73*, 1602–1611.
(6) Lan, T.; McLaughlin, L. W. *J. Am. Chem. Soc.* **2000**, *122*, 6512–6513.
(7) Lan, T.; McLaughlin, L. W. *Biochemistry* **2001**, *40*, 968–976.
(8) Sun, Z.; McLaughlin, L. W. *Biopolymers* **2007**, *87*, 183–195.
(9) Siegfried, N. A.; Kierzek, R.; Bevilacqua, P. C. *J. Am. Chem. Soc.* **2010**, *132*, 5342–5344.
(10) Irani, R. J.; Santalucia, J. *Nucleosides, Nucleotides Nucleic Acids* **2002**, *21*, 737–751.
(11) Arico, J.; Calhoun, A. K.; Salandria, K.; McLaughlin, L. W. *Org. Lett.* **2010**, *12*, 120–122.

Chapter 4: Fluorescent Labeling in Minor Groove of A:T Base Pair

4.1 Fluorescent Nucleic Acids

Labeled nucleic acids offer a mechanism to track nucleic acids in biological and laboratory environments. Labeling techniques include radioactive ^{32}P phosphorylation, intercalation, and fluorescent tags. Radiation and intercalation pose health dangers due to their high toxicities; fluorescence labeling provides a safer means to DNA detection. Fluorescent labeling is an ideal tool to track biological processes in the cell. Small molecules and proteins can be covalently attached to biomolecules of interest (DNA, RNA, antibodies) and their interactions within their environment can be monitored by fluorescent properties. Potential uses include FRET (Förster Resonance Energy Transfer)[1] and anisotropy[2] studies of protein and nucleic acid interactions. Inherent fluorescence of nucleic acids are limited.[3] Two modes exist for labeling DNA: covalent and non-covalent bonds. Intercalators such as ethidium bromide, slide in between base pairs, but do not form a covalent bond to the duplex. The hydrophobic environment of the helix core causes ethidium to fluoresce, allowing binding to DNA to be monitored. Radiolabeling of nucleic acids can be accomplished by kinase labeling with γ-^{32}P-ATP. Fluorophores are covalently bonded to the molecule of interest by either naturally or synthetically incorporated labeling sites

Inherent sites for labeling are found on all three units of a nucleotide monomer, making site-specific labeling difficult to control. Use of synthetically modified nucleotides is a simpler approach to unique labeling sites in DNA.[4] Amines, thiols,

azides, and carboxy-groups can be incorporated into nucleotides that react specifically with chemically activated fluorophores. Previous work has shown modification of the phosphodiester backbone,[5,6] sugar moiety,[7] and the heterocycles.[8] Labeling of the heterocycles has only been demonstrated in the major groove to date, the minor groove of DNA is a novel site for fluorescence labeling. Our work with minor groove modifications allows for direct incorporation of labeling sites of the 3-position of 2′-deoxyadenosine monomers.

Our work with tetramethylsuccinimide has offered synthetic access to a variety nucleosides containing functional groups of different steric sizes off the 3-position of 2′-deoxyadenosine. We are now able to install free amines into the minor groove of DNA, which can be used for labeling. Expanding on this work should allow us to install free thiols into the minor groove as well. These groups can be reacted with fluorophores prior to or post-DNA synthesis to function as a new reporter group on nucleic acids.

4.2 Attempts at Synthesis of New Minor Groove Labeling Sites

An abundance of fluorophores consisting of mixed and matched reporter groups and amine- or thiol-specific functional groups are commercially available. Our goal included tethering amine and thiol functionality (**Figure 1**) into the minor groove to demonstrate how reacting these sites with fluorophores can affect duplex DNA. Incorporation of one modified residue into DNA grants access to site-specific labeling within the minor groove. Our target DNA sequence, d(GGCC AA$\boldsymbol{A}^{\boldsymbol{CH2R}}$A CGCC) offers a single labeling site within duplex DNA. The synthesis of damc$_3$A phosphoramidite and its incorporation into DNA have been outlined in Chapter 3.

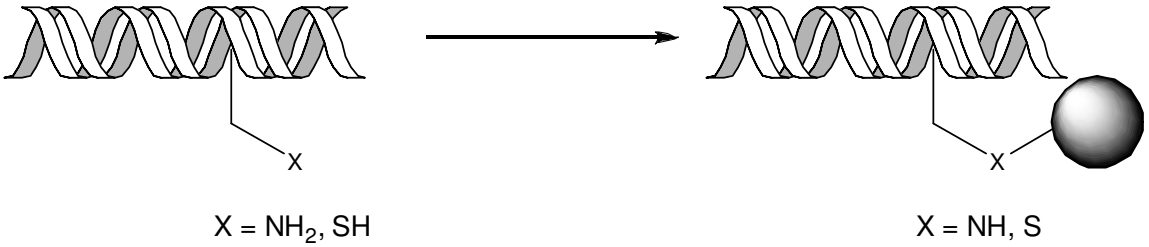
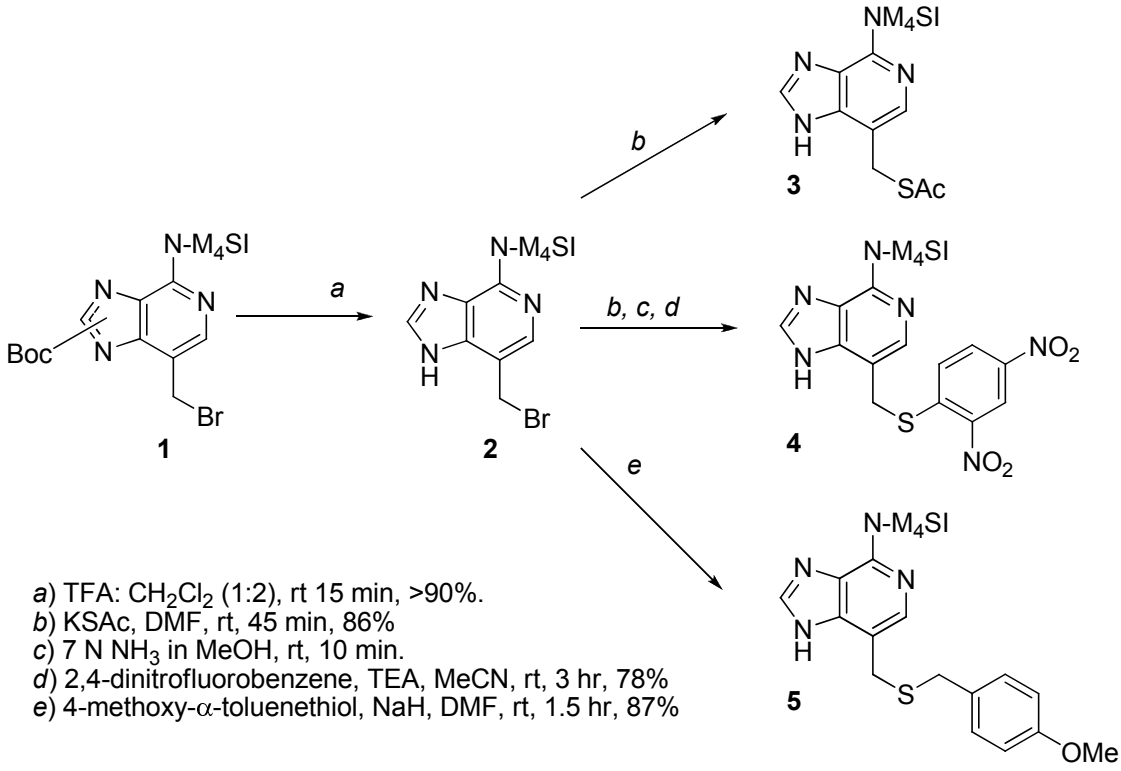


Figure 1. Labeling using amino- and thio-methyl modified 3-deaza-2′-deoxyadenosine

Synthesis of the thiomethyl protected phosphoramidite involved much trial and error as selecting an appropriate protecting group proved to be rather difficult. A series of different groups and approaches were taken, but one theme remained consistent: slight heating of this reaction in the presence of thiolates led to poor yields and multiple side products. The optimal route involves removal of the Boc group prior to thiolate displacement with standard TFA:CH_2Cl_2 deprotection (**Scheme 1**) to avoid heating

Scheme 1. Synthesis of Potential Thio-Protected Nucleobases.



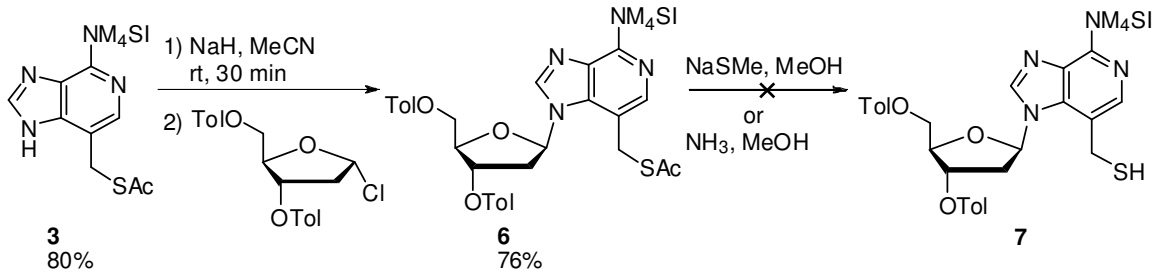
during the displacement reactions. Potential protecting groups that would be stable throughout DNA solid-phase synthesis were acetate thioester, 2,4-dinitrophenyl thioether, and *p*-methoxybenzyl thioether. The acetate group is labile to standard ammonium hydroxide treatment, while the 2,4-dinitrophenyl thioether requires extra treatment with 2-mercaptoethanol (BME) to remove the protecting group.[9] *P*-methoxybenzyl thioethers are used to protect cysteine residues during peptide solid phase synthesis.[10] These groups are removed using mercuric (II) acetate in TFA in the presence of anisole.

The thioacetate group is installed by displacement of bromo-nucleobase **2** with KSAc. The acetate can be removed with methanolic ammonia resulting in a free thiol, which can then be reacted with 2,4-dinitrofluorobenzene to give **4**. The next group

explored was installed by displacement of bromo-nucleobase **2** with 4-methoxy-α-toluenethiol to give **5**.

The first group examined was thioacetate. The bromo-nucleobase reacted with KSAc to afford thioester **3** in 86% yield. Glycosylation of heterocycle **3** yielded thio-protected nucleoside **10** in 76% yield (**Scheme 2**). Selective deprotection of the toluoyl esters could not be accomplished in the presence of the thioester. Sodium thiomethoxide in MeOH has been used to deprotect thioesters in the presence of *O*-esters,[11] but treatment of **6** with these conditions resulted in unwanted deprotection of the toluoyl groups after 30 minutes. Removal of both thioacetate and the toluoyl esters with 7 N NH_3 in MeOH resulted in a crude reaction mixture that was difficult to purify as well as low yielding. Discriminating between *O*-esters and *S*-esters proved challenging, so new methods for protecting the thiol functionality were explored.

Scheme 2. Attempts at Selective Deprotection of Thioester in the Presences of *O*-esters.



The thioacetate displacement could be used to successfully install thio-functionality before glycosylation. Deprotection of nucleobase **3** with 7 N NH_3 afforded the free thiol, which could then be protected with 2,4-dinitrophenyl (DNP) thioether to give **4**. Selective protection of the thioether proved rather difficult as the imidazole

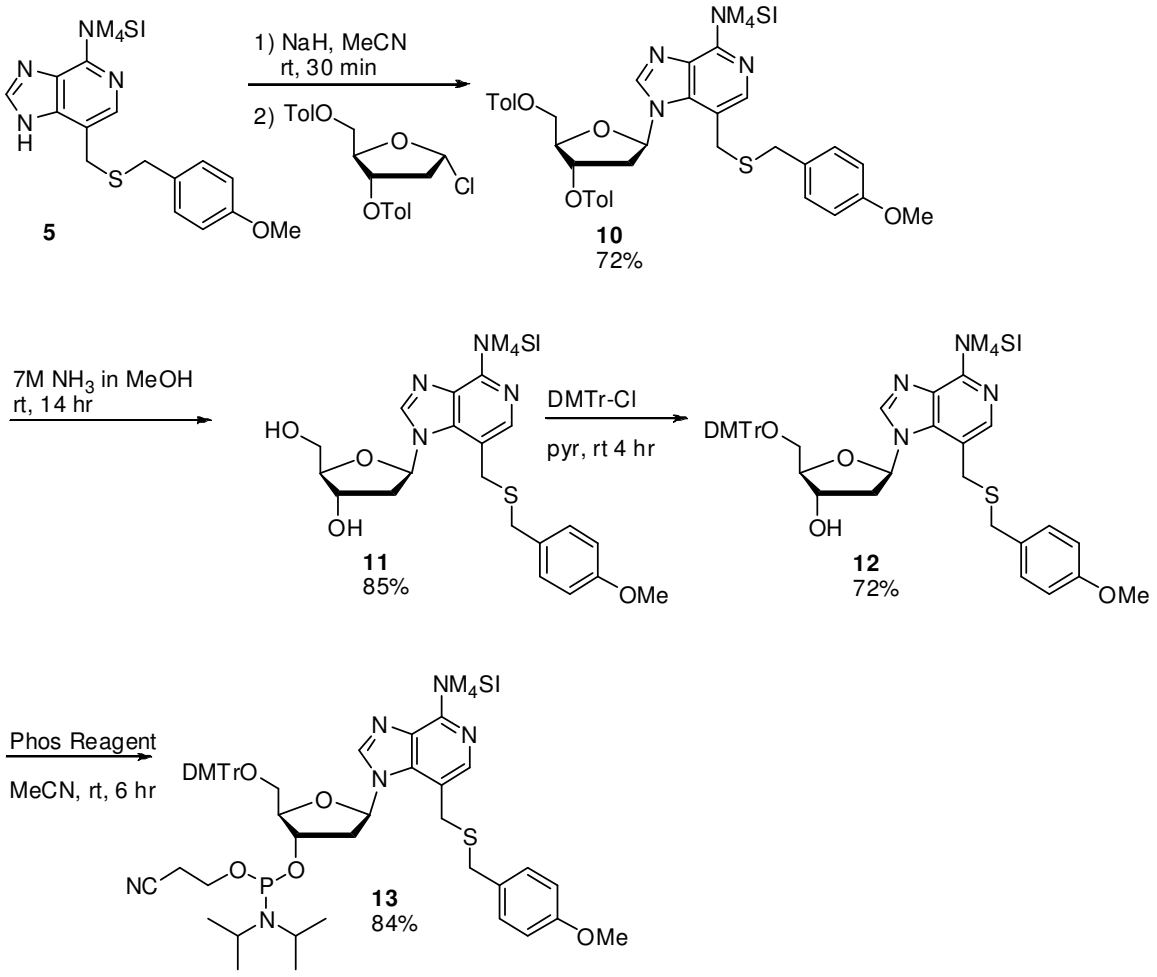
moiety of the heterocycle is somewhat nucleophilic and reactive towards the protecting group. The reaction resulted in multiple products, including the desired product **4**, but also β-N9 and β-N7 DNP-nucleosides.

We next focused our efforts on *p*-methoxybenzyl thioether since it is commonly used as a protecting group for peptide solid-phase synthesis on cysteine residues. Displacement of the bromide with 4-methoxy-α-toluenethiolate successfully installed the thiol functionality and protection simultaneously to give **5**. This analogue was successfully glycosylated and carried throughout the synthesis to phosphoramidite **13** (**Scheme 3**). The phosphoramidite was incorporated into DNA with yields similar to those of commercially available phosphoramidites.

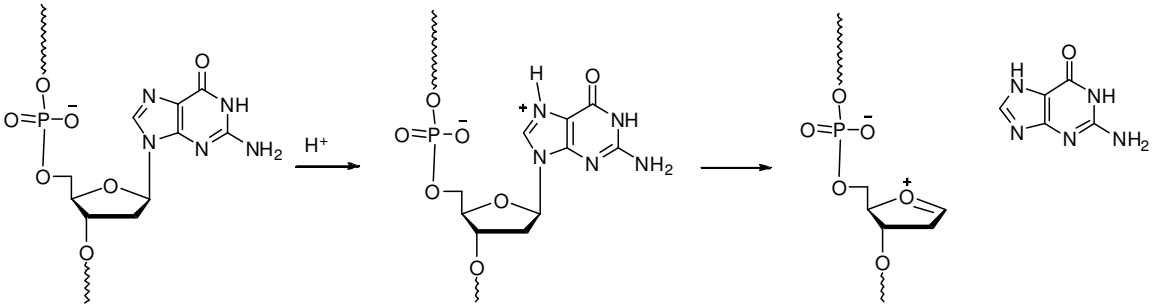
Removal of the protecting group post-DNA synthesis was fruitless. Standard deprotection conditions call for mercuric (II) acetate in trifluoroacetic acid with anisole.[12, 13] Treatment of DNA with TFA led to depurination of the purine nucleosides (**Scheme 4**). It was clear that TFA was too strong of an acid to use with DNA. We sought to use a weaker acid, such as acetic acid or even water, as the solvent with either mercuric (II) acetate or mercuric (II) trifluoroacetate. These attempts were also unsuccessful.

The chosen protecting group had to be stable to ammonia in methanol conditions to remove the toluoyl groups, but labile to ammonium hydroxide deprotection post-DNA synthesis. Cyanoethyl thioethers are commonly used to protect 6-thio-2′-deoxyguanosine and 4-thiouracil phosphoramidites. They are used throughout DNA synthesis and removed with NaOH containing a small amount of NaSH,[14] so this group seemed best for

Scheme 3. Synthesis of p-Methoxy-Benzylthioether Phosphoramidite (**9**)

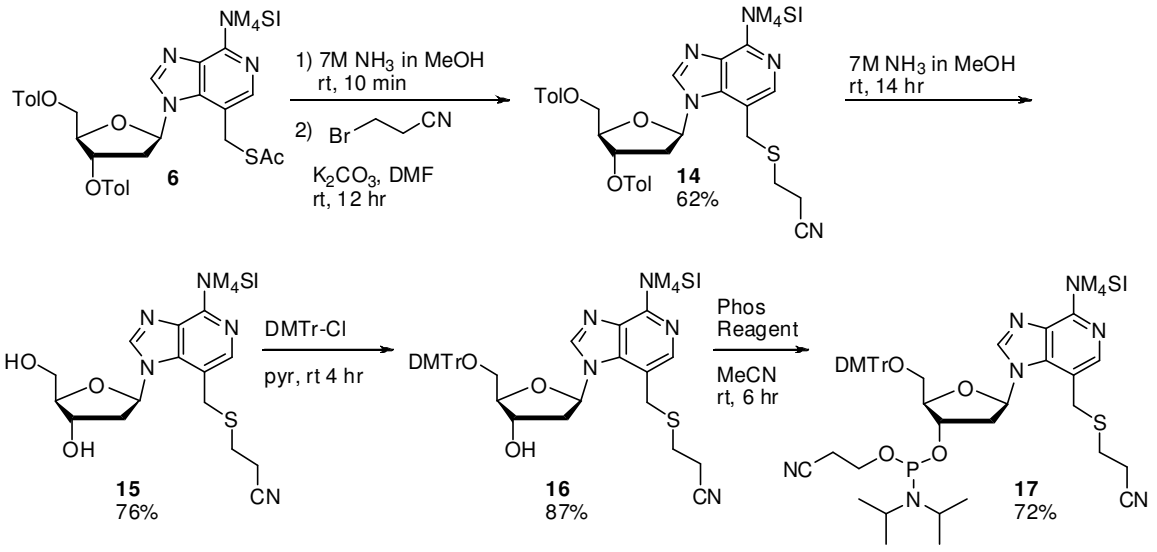


Scheme 4. Depurination of Purine Residues in DNA with acid.



our purposes. Displacement of the bromide with potassium thioacetate afforded **3** in acceptable yield and was stable to glycosylation conditions, yielding 76% of the desired isomer (**Scheme 2**). The cyanoethyl protecting group could be installed in one pot by treatment of **6** with NH_3 in methanol for 10 minutes (**Scheme 5**). This did lead to small amounts of toluoyl ester deprotection, however it seemed the most direct route to deprotecting the thioester. The free thiol can be treated with 3-bromopropionitrile and potassium carbonate in DMF to yield protected thioether **14** in 62% yield. It should be noted that appropriate monitoring of this reaction must take place. Long reaction times in DMF lead to isomerization of the β-N9 nucleoside. The toluoyl groups could then be removed cleanly in methanolic ammonia to give **15** in 76% yield with no thioether deprotection detected. Elaboration to dtc_3A phosphoramidite by DMT-protection followed by phosphitylation resulted in phosphoramidite **17** in 63% overall yield over 2 steps. DNA synthesis with **17** incorporated a single dtc_3A residue.

Scheme 5. Synthesis of 2-cyanoethyl thioether Phosphoramidite **17**



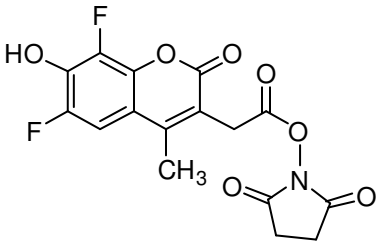
Deprotection of the 2-cyanoethyl thioether proved rather difficult. The reaction proceeds via β-elimination to produce the free thiol and cyanoethene. All references to date show this protecting group utilized on aromatic thiols.[15-17] Due to the poor leaving group character of the benzyllic thiol, bases such as NaOH are not strong enough to remove the proton α- to the nitrile group. The pKa of this proton is typically around 25, so much stronger bases like sodium amide must be used to effectively deprotonate the α-position and allow β-elimination to take place. Optimally, the chosen protecting group will be labile to standard ammonium hydroxide treatment. Prospects on these groups and extensions of the work above are discussed in section 4.4

4.3 Fluorescent Labeling of DNA Minor Groove

The amc_3A minor-groove labeling site was reacted with specific probes after standard deprotection and purification of the synthesized oligos. Typical conditions involved reaction of 50 nmol of DNA with 20-fold molar excess of the label in a solution of 1:10 DMSO:Buffer (30 mM $NaHCO_3$, pH 8.45). The reaction was incubated overnight at room temperature to provide the labeled oligo, which could be purified by HPLC (**Figure 2**, next page). HPLC traces of labeling reactions followed the same general pattern: elution of non-labeled DNA, followed by fluorescent-labeling DNA, and finally elution of unreacted fluorophores. The desired peak was collected, concentrated, and desalted using a Sep-Pak C_{18} cartridge (Waters, Milford, MA). Melting temperature data was then obtained by hybridizing the labeled oligo to its complement.

We chose two labels differing in fluorescent tags and length of the linker (**Figure 3**). DNA containing the $damc_3A$ analogue readily reacted with both Marina Blue and Fluorescein-5-EX succinimidyl esters in 87% yield.

Marina Blue Succinimidyl Ester



Fluorescein-5-EX Succinimidyl Ester

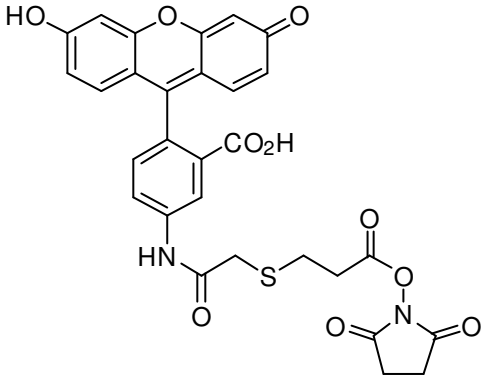


Figure 3. Amine Reactive Fluorophores

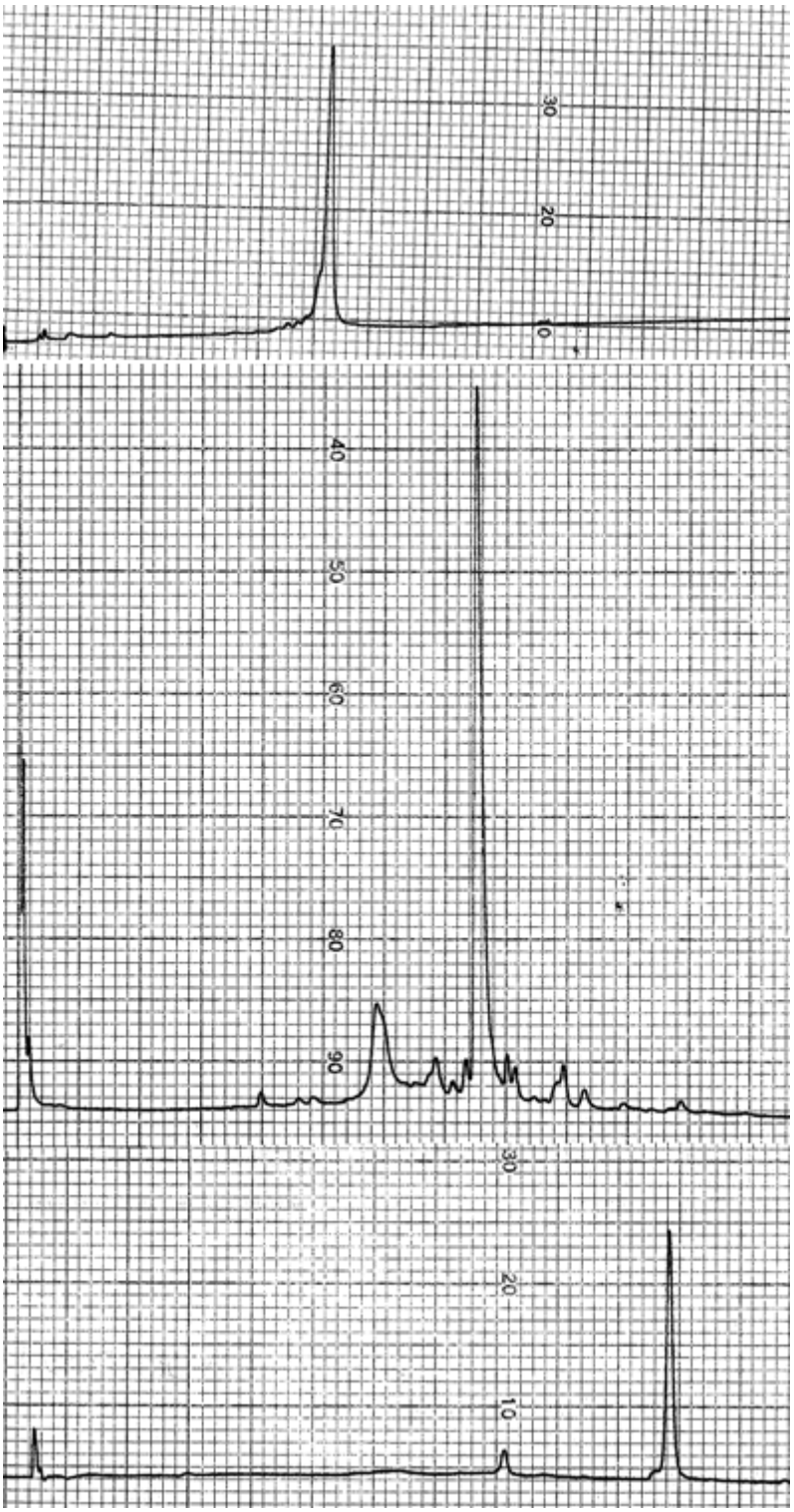


Figure 2. HPLC traces of unmodified amc_3A DNA (top), labeling reaction with Fluorescein-5-EX (middle) and pure Fluorescein-5-EX succinimidyl ester (bottom).

The results of fluorescent labeling in the minor groove are shown in **Table 1**. Labeling with Marina Blue adds 2.9 ^{o}C to the amino-control sequence and 2.4 ^{o}C to the native sequence, suggesting this fluorophores acts to stabilize the helix. This can be attributed to dynamic intercalation; intercalators are positively charged and this fluorophore is neutral. It is thought that the fluorophore slides in and out of the base pairs, offering some additional stability, but because it is dynamic, it is not technically an intercalator. Amino-labeling with fluorescein-5-EX decreases the melting temperature, T_M, by 4.1 ^{o}C relative to the amino-substituted oligo. This may be attributed to the additional length of linker or the larger size of the fluorophores as compared to Marina Blue.

Table 1. Melting Temperatures of Minor Groove Labeled DNA

Sequence	Fluorophore	T_M [†]
d(CCGG AAAA CGCC)	---	63.4
d(CCGG AA***A***^{CH2NH2}A CGCC)	---	62.9
d(CCGG AA***A***$^{CH2NH\text{-}MB}$A CGCC)	Marina Blue	65.8
d(CCGG AA***A***$^{CH2NH\text{-}F5EX}$A CGCC)	Fluorescein-5-EX	58.8

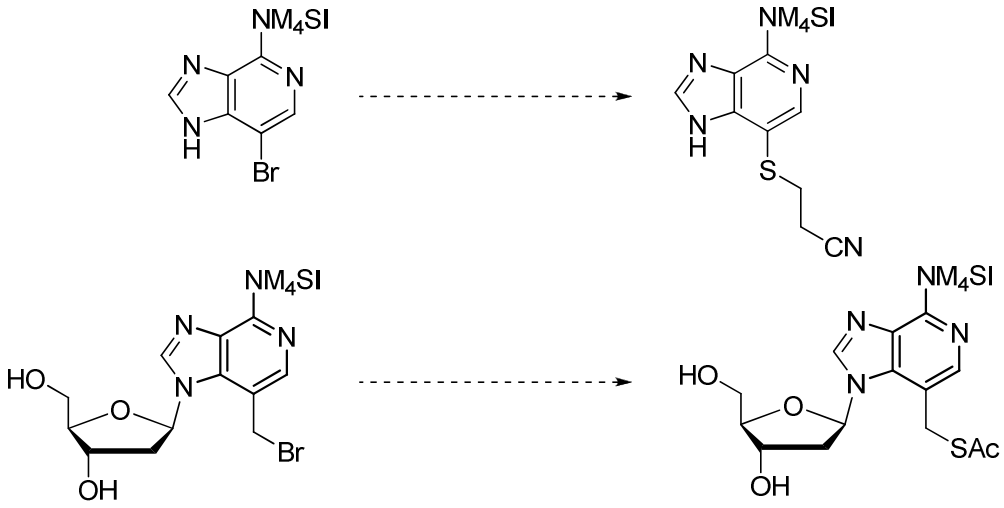
[†] 5 uM concentration of duplex, 20 mM Phosphate Buffer, pH 7, 1 M NaCl.

4.4 Conclusions and Future Work

The work described throughout this thesis details how minor groove modifications can be used to study the minor groove of DNA. Tetramethylsuccinimide is a valuable asset to accessing synthetic modifications on the 3-position of 2′-deoxyadenosine prior to glycosylations. Previous to this work, few modifications in the minor groove of A-T base pairs existed, especially those presenting somewhat sterically large substituents.

The modifications presented in Chapter 2, Figure 2 were designed to study how small molecules such as water and ammonium cations bind to A-tract DNA in the minor groove. The results presented in Chapter 3 show that these molecules contribute somewhat to the structure and stability of duplex DNA. Further study of molecules binding in the minor groove arises from aminomethyl ($damc_3A$) and thiomethyl (dtc_3A). These two analogues can be treated with amine or thiol specific fluorophores to give rise to a new site for labeling within DNA. Minor groove labeling has not been shown before, but can be used to study how molecules, such as proteins and drugs, bind to the region of duplex DNA. Currently, need protecting groups and deprotection methods are being explored for tmc_3A incorporation into DNA. Possible deprotection with sodium amide or other NR_2^- bases may function to deprotonate the α-proton on the 2-cyanoethyl protecting group, but these conditions are rather dangerous and difficult to work with safely.

Scheme 6. Proposed Steps towards new tmc_3A Analogues.



New approaches to successful installation of tmc_3A into DNA include changing the structure or synthetic route of the thiomethyl derivative (**Scheme 6**). If the methylene unit is removed the the thiol moiety is directly attached to the nucleobase, this should activate the leaving group enough to allow for deprotection to occur under standard conditions. A reaction that has not been explored is coupling the 3-deaza-3-methylbromopurine to the sugar. If this reaction is successful, the toluoyls can be removed prior to thioacetate displacement. This would allow direct elaboration to the phosphoramidite by DMT protection and phosphitylation.

On-going and future work for this project also includes expansion of protecting and directing groups for purine glycosylations as shown in Chapter 2,4. The ultimate goal is two-fold: i) to increase the aliphatic character of the protecting group as to make these protected nucleobases more soluble in non-polar solvents and ii) increase the steric size of the substituents of the succinimide group to more effectively block the N7-position during glycosylation. Meetings these goals would increase the overall yield of desired β-

N9 isomer and give way to more commercially available modified nucleosides at modest pricing.

In order to expand our studies on how small molecules bind the minor groove, current work involves expanding the linker of the functional groups by one methylene unit (**Figure 4**). It remains uncertain whether one methylene unit is an adequate linker to permit the hydroxyl group to reach across the helix to hydrogen bond with the O^2-carbonyl of thymidine, one base pair displaced. Results from these studies will reveal in more detail the required length of linker for optimal hydrogen bonding.

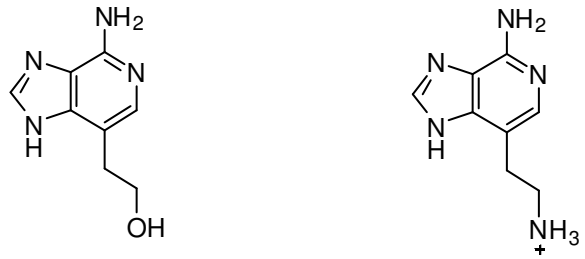


Figure 4. Extension of Minor Groove Modification by One Methylene Unit.

Factors that stabilize or disturb duplex formation are important in the design of molecules targeted towards DNA. Our research demonstrates that water and ammonium binding play a role in stabilizing the thermodynamic parameters and structural integrity of duplex DNA. The key to synthesizing these important analogues is tetramethylsuccinimide as a directing and protecting group. The results from these studies have enabled researchers to gain insight into the minor groove in ways that was previously not established.

4.5 Experimental

***N*-(7-(acetylthiomethyl)-1*H*-imidazo[4,5-*c*]pyridin-4-yl)-2,2,3,3-tetramethylsuccinimide (3).** The Boc group wass first removed by dissolving bromo-nucleobase **1** (1.58 g, 3.42 mmol) in CH_2Cl_2 (20 mL) and adding TFA (10 mL). After 10 min, TLC (1:1 Hex:EtOAc) indicated complete consumption of starting material. Water (10 mL) was added and the solution was transferred to a separatory funnel. After vigorous shaking, the organic layer was separated and washed with sat. $NaHCO_3$ (2 x 10 mL). The organic layer was dried over Na_2SO_4, filtered and concentrated to yield 938 mg (77%) of the deprotected, bromo-substituted nucleobase **2**. This nucleobase (500 mg, 1.37 mmol) was dissolved in DMF and KSAc was added (235 mg, 2.06 mmol). The solution stirred for 1 hr at room temperature upon which time TLC indicated complete conversion to **1**. DMF was removed *in vacuo* and the crude product was applied to silica gel chromatography. Elution with 5% MeOH in DCM produces 424 mg (86%) of the title compound as a white solid. ^{1}H NMR (500 MHz, acetone-*d6*) δ 8.31 (s, 1H), 8.30 (s, 1H), 4.48 (s, 2H), 2.36 (s, 3H), 1.32 (s, 12H). ^{13}C NMR (126 MHz, acetone-*d6*) δ 190.99, 180.90, 143.81, 143.68, 141.15, 138.16, 137.03, 119.53, 47.55, 26.77, 20.83, 20.66 ppm. HRMS (ESI-TOF) calcd for $C_{17}H_{21}N_4O_3S$ $[M+H]^+$ 361.1334, found 361.1335.

1-[2-deoxy-3,5-di-*O*-(*p*-toluoyl)-β-D-*erythro*-pentofuranosyl]-4-(3,3,4,4-tetramethyl-2,5-dioxopyrrolidin-1-yl)-7-acetylthiomethyl-1*H*-imidazo[4,5-*c*]pyridine (10). To a flask containing **3** (83 mg, 0.230 mmol) and NaH (17 mg, 0.345 mmol, 50% dispersion in oil) was added 5 mL of MeCN. The mixture was allowed to stir for 45 min. The α-chloroprotected sugar (161 mg, 0.414 mmol) was added and stirring continued for

2 h. The reaction mixture was filtered and volatiles were removed *in vacuo* and the resulting yellow foam was purified by flash chromatography eluting with 1:1 Hexanes:EtOAc to afford 125 mg (76%) of **10** as a white foam. ^{1}H NMR (500 MHz, acetone-*d*6) δ 8.61 (s, 1H), 8.33 (s, 1H), 8.01 (d, *J* = 9 Hz, 2H), 7.86 (d, *J* = 8 Hz, 2H), 7.34 (d, *J* = 8.5 Hz, 2H), 7.26 (d, *J* = 8 Hz, 2H), 6.89 (dd, *J* = 5, 3, 5.5 Hz, 1H), 5.82 (m, 1H), 4.80 (d, *J* = 14.5, 1H), 4.73 (m, 1H), 4.63 (d, *J* = 4.5 Hz, 2H), 4.60 (d, *J* = 14.5 Hz, 1H), 3.23 (m, 1H), 3.11 (m, 1H), 2.39 (s, 3H), 2.38 (s, 3H), 2.35 (s, 3H), 1.32 (d, *J* = 27.5 Hz, 12H) ppm. ^{13}C NMR (126 MHz, acetone-*d6*) δ 193.51, 180.59, 165.49, 144.34, 143.92, 143.51, 143.29, 139.07, 138.65, 138.05, 129.75, 129.72, 129.66, 129.52, 129.26, 129.24, 127.12, 127.03, 118.86, 85.63, 82.75, 74.79, 63.99, 47.62, 37.54, 27.29, 21.01, 20.82, 20.75, 20.73, 20.71, 20.55 ppm. HRMS (ESI-TOF) calcd for $C_{38}H_{40}N_4O_8SNa^+$ [M + Na]$^+$ 735.2465, found: 735.2465.

1-[2-deoxy-3,5-di-*O*-(*p*-toluoyl)-β-D-*erythro*-pentofuranosyl]-4-(3,3,4,4-tetramethyl-

2,5-dioxopyrrolidin-1-yl)-7-(2-cyanoethyl)thiomethyl-1*H*-imidazo[4,5-*c*]pyridine (12). To a flask containing **10** (50 mg, 0.070 mmol) was added MeOH containing 7 M NH_3 (1 mL). The flask was tightly stoppered and the mixture allowed to stir for 10 min. TLC indicated complete consumption of **10** and the solution was concentrated *in vacuo*. The free thiol was then protected by dissolving the product in DMF (700 uL) and addition of 3-bromopropionitrile (6 uL, 0.70 mmol) in the presence of K_2CO_3 (19 mg, 0.14mmol). The solution stirred for 10 hr and was then concentrated. Flash column chromatography (9:1 DCM:MeOH) gave 31 mg (62%) of pure **12**. ^{1}H

NMR (500 MHz, acetone-*d6*) δ 8.61 (s, 1H), 8.31 (s, 1H), 8.04 (d, *J* = 5 Hz, 2H), 7.89 (d, *J* = 7.5 Hz, 2H), 7.40 (d, *J* = 8.5 Hz, 2H), 7.30 (d, *J* = 5.5 Hz, 2H), 7.02 (dd, *J* = 13.5, 5.5 Hz, 1H), 5.89 (q, *J* = 12, 3 Hz, 1H), 4.81 (q, *J* = 12.5, 2.5 Hz, 1H), 4.66 (d, *J* = 5 Hz, 2H), 4.60 (d, *J* = 13.5 Hz, 1H), 4.37 (d, *J* = 13.5 Hz, 1H), 3.28-3.25 (m, 1H), 3.21-3.18 (m, 1H), 2.95 (m, 4H), 2.42 (s, 3H), 2.39 (s, 3H), 1.36 (d, *J* = 26.5 Hz, 12H). ^{13}C NMR (126 MHz, acetone-*d6*) δ 165.61, 165.44, 144.32, 143.90, 143.24, 142.93, 139.28, 138.80, 138.21, 129.66, 129.48, 129.28, 129.23, 127.10, 127.01, 119.17, 118.57, 86.00, 82.63, 74.76, 63.92, 47.60, 37.63, 27.08, 20.99, 20.71, 20.66, 20.51, 17.98. HRMS (ESI-TOF) calcd for $C_{39}H_{42}N_5O_7S$ $[M+H]^+$ 724.2805, found 724.2778.

1-[2-deoxy-β-D-*erythro*-pentofuranosyl]-4-(3,3,4,4-tetramethyl-2,5-dioxopyrrolidin-

1-yl)-7-(2-cyanoethyl)thiomethyl-1*H*-imidazo[4,5-*c*]pyridine (13). To a flask containing **12** (50 mg, 0.069 mmol) was added MeOH containing 7 M NH_3 (1.5 mL). The flask was tightly stoppered and the mixture allowed to stir for 14 h. Volatiles were removed *in vacuo* and the residue purified by flash chromatography eluting with 8% MeOH in DCM to afford **13** as a white foam (26 mg, 76%). ^{1}H NMR (acetone-*d6*, 500 MHz) δ 8.63 (s, 1H), 8.28 (s, 1H), 6.79 (t, *J* = 6.5 Hz, 1H), 4.69 (q, *J* = 9.5, 2 Hz, 1H), 4.51 (d, *J* = 13 Hz, 1H), 4.30 (d, *J* = 13.5 Hz, 1H), 4.11 (q, *J* = 7, 3.5 Hz 1H), 3.77 (m, 2H), 2.91-2.85 (m, 5H), 2.73 (m, 1H), 1.36 (d, *J* = 26 Hz, 12H) ppm. ^{13}C NMR (acetone-*d*6, 125 MHz) δ 181.21, 143.89, 141.63, 139.37, 138.42, 137.76, 122.33, 88.00, 85.99,

70.52, 61.53, 50.14, 47.58, 31.37, 21.05, 20.89 ppm; ^{13}C NMR (126 MHz, acetone-*d6*) δ 180.89, 143.86, 142.47, 138.99, 138.68, 138.20, 119.14, 118.65, 88.39, 86.01, 71.01, 61.76, 47.57, 27.11, 20.91, 20.64, 17.95. HRMS (ESI-TOF) calcd for $C_{23}H_{30}N_5O_5S$ $[M+H]^+$ 488.1968, found 488.1957.

1-[5'-*O*-(4,4'-dimethoxytrityl)-2-deoxy-β-D-*erythro*-pentofuranosyl]-4-(3,3,4,4-tetramethyl-2,5-dioxopyrrolidin-1-yl)-7-(2-cyanoethyl)thiomethyl-1*H*-imidazo[4,5-*c*]pyridine (14): To a flask containing **13** (74 mg, 0.152 mmol) and DMTr-Cl (61 mg, 0.182 mmol) was added pyridine (3 mL). The solution was allowed to stir for 4 hr. TLC (95:5 DCM:MeOH) indicated complete consumption of starting material. The reaction solution was concentrated *in vacuo* and applied to silica gel. Column Chromatography with 1% TEA in EtOAc afforded 104 mg (87%) of the DMT-protected nucleoside **14**. ^{1}H NMR (acetone-*d6*, 500 MHz) δ 8.47 (s, 1H), 8.33 (s, 1H), 7.39-6.77 (m, 14H), 4.71 (q, *J* = 4.5, 3.5 Hz, 1H), 4.63 (d, *J* = 13.5 Hz, 1H), 4.34 (d, *J* = 13.5 Hz, 1H), 4.26 (q, *J* = 4, 3 Hz, 1H), 4.03 (m, 1H), 3.74 (s, 6H), 3.30 (d, *J* = 4 Hz, 2H), 3.00 (m, 1H), 2.92 (m, 4H), 2.77 (m, 1H), 1.36 (d, *J* = 27 Hz, 12H) ppm.^{13}C NMR (acetone-*d6*, 126 MHz) δ 180.89, 158.67, 144.99, 143.31, 142.60, 139.14, 138.88, 138.41, 135.70, 130.00, 129.96, 128.00, 127.69, 126.66, 119.33, 118.65, 113.00, 86.65, 86.10, 85.70, 71.02, 63.98, 63.55, 54.61, 47.59, 39.90, 27.08, 21.09, 20.83, 20.71, 20.46, 19.90, 18.86, 17.94, 13.62, 13.07 ppm. HRMS (MALDI-TOF) calcd for $C_{45}H_{50}N_6O_8SNa$ $[M+Na]^+$ 812.3088, found 812.3103.

3-deaza-3-thiomethyl-2'-deoxyadenosine phosphoramidite 15. To a flask containing 5'-*O*-DMT-protected **14** (72 mg, 0.091 mmol) dissolved in MeCN (1.8 mL) at 0 ^{o}C was added tetrazole (303 uL of 0.45 M solution in MeCN, 0.136 mmol) followed by 2-cyanoethyl-*N*,*N*,*N*,*N*-tetraisopropylphosphane (87 μL, 0.274 mmol) dropwise. The mixture was allowed to warm up and stir at rt for 6 h, after which TLC indicated that the reaction was complete. Methanol (1 mL) was added and stirring continued for 10 min. The reaction was concentrated *in vacuo* and applied to silica gel column chromatography (gradient of 40% Hexanes, 1% TEA, 59% EtOAc to 100% EtOAc). Precipitation by slow addition of the crude mixture dissolved in minimal DCM to vigorously stirring hexanes followed by cooling at 0 °C for 30 min afforded **6** as a white solid (65 mg, 72%, 1:1 mixture of diasteriomers). ^{31}P NMR (acetone-*d6*) δ 148.66, 148.50.

DNA Synthesis.

The sequence used for this study is the same as those described in Chapter 3, d(CCGG AAAA CGCC). A single modification of the third A with dtmc$_3$A afforded a non-self complementary oligomer with one specific site used for fluorescence labeling. The modified 12-mers was prepared by solid-phase DNA synthesis and deprotected using standard protocols. The analogue nucleotide **15** could be incorporated into DNA strands with essentially the same coupling efficiency as the common nucleotides using the standard wait time of *2 min* during the coupling cycle. The oligo was deprotected and cleaved from the CPG beads by 16 h treatment with concentrated ammonia at 55 °C. The solutions were then decanted and evaporated by speedvac, then redissolved in 4-6 mL H_2O for HPLC purification.

Initial purification the modified oligonucleotide was accomplished by HPLC (Oligo R3 reverse-phase C18 column, trityl on), starting with 100% A using a linear gradient from 0 to 50% B over 28 min [A: 1 M TEAA (pH 7) in 5% acetonitrile; B: 1 M TEAA (pH 7) in 70% acetonitrile]. The DMT-protected 12-mer had retention times of about 15 minutes. The corresponding peak was collected and then reduced in volume and detritylated.

Detritylation in 80% acetic acid for 60 min at 0 °C, followed by desalting (Sephadex G-10) and lyophilization gave the $dtmc_3A$ modified oligomer as a white powder. Mass Spectral Analysis (MALDI-TOF) showed a peak corresponding to the $dtmc_3A$ oligomer with the 2-cyanoethyl protecting group still present. Further treatment of this oligo in 1 M NaOH at 50 ^oC was monitored by HPLC until the retention time of the peak shifted from 12.8 minutes to 10.2 minutes (Waters C_{18} Column, 0 to 100% B over 60 min. A: 50 mM TEAA, pH 7, 5% Acetonitrile; B: 50 mM TEAA, pH 7, 70% Acetonitrile).

Fluorescent Labeling

To begin, 50 nmol of singly modified $damc_3A$ or $dtmc_3A$ oligomers were lyophilized. The label was prepared by dissolving 1 mg in 100 uL of DMSO. The lyophilized oligo was dissolved in 90 uL buffer (30 mM $NaHCO_3$, pH 8.45 for $damc_3A$ and 40 mM Phosphate Buffer, pH 9.2 for dtc_3A) and 10 uL of the label (about 10-20 molar equivalents) was added to the DNA solution. The $damc_3A$ containing reactions were allowed to sit overnight at room temperature on a shaker plate, while the $dtmc_3A$

containing reactions were allowed to react for 3 hr at room temperature.

The newly labeled oligomers were purified from unreacted oligo and label via HPLC (Waters C_{18} Column, 0 to 100% B over 60 min. A: 50 mM TEAA, pH 7, 5% Acetonitrile; B: 50 mM TEAA, pH 7, 70% Acetonitrile). Unmodified oligos had retention times from 10-15 minutes, while labeled oligos had retention times of 15-18 minutes. The unreacted fluorophores eluted last between 22 and 25 minutes. The desired peak was collected and any acetonitrile was removed *in vacuo*. The solution was then desalted via C_{18} Sep Pak (Waters) and lyophilized.

Thermal Denaturation Studies

Thermal denaturation studies were performed in solutions of 20 mM NaH_2PO_4 (pH 7.0) and 1 M NaCl with the concentrations of duplex at 5 μM). Absorbance and temperature values were measured with an AVIV 14DS UV-visible spectrophotometer equipped with digital temperature control. The temperature of the cell compartment was increased in 1.0 °C/min steps (from 4 to 95 °C), and when equilibrium was reached, temperature and absorbance data were collected every 1 ^{o}C. The raw data was exported to the software application Microcal Origin (OriginLab Corporation, Northampton, MA) for determination of the melting temperature (T_m) and free energy (ΔG) of each modified duplex. First, the T_m plot (absorbance vs temperature) was converted to an alpha (α) plot (fraction of molecules paired vs temperature) using the equation:

$$\alpha = \frac{A - A_s}{A_d - A_s} \quad (1)$$

where A is the absorbance at a given temperature, and A_s and A_d are the absorbance values for the single- and double-stranded states, respectively. The parameter α represents the fraction of molecules paired; the T_m is extrapolated when $\alpha = 0.5$. In order to obtain sufficient approximations for A_s and A_d, the upper and lower baselines of the melting curve were fitted using linear least-squares fits. The thermodynamic parameters ΔH, ΔS, and ΔG were determined using a van't Hoff plot (ln K vs $1/T$). The value of K was calculated at each temperature from the α plot using the following equation:

$$K = \frac{\alpha}{2(1-\alpha)^2 C_t} \quad (2)$$

where C_t is the total strand concentration. The thermodynamic parameters can be derived directly from the linear least-squares fit in which the slope is $-\Delta H^o/R$ and the intercept is $\Delta S^o/R$. The free energy change (ΔG^o) at any temperature can be determined by using the following relationship:

$$\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ \quad (3)$$

References

1) Joseph R. Lakowicz, "Principles of Fluorescence Spectroscopy", Plenum Publishing Corporation, 2nd edition (July 1, 1999).

2) Heyduk, T., Ma, Y., Tang, H., Ebright, R. H. *Methods Enzymol.*, **1996**, *274*, 492-503.

3) Kricka, L. J. and Fortina, P. *Clinical Chem.,* **2009**, *55*, 670-683.

4) Ozaki, H. and McLaughlin, L. W. *Nucleic Acids Res.,* **1992**, *20(19),* 5205-5214.

5) Fidanza, J. A. and McLaughlin, L. W. *J. Am. Chem. Soc.*, **1989**, *111(1),* 9117-9119.

6) Fidanza, J. A., Ozaki, H., and McLaughlin, L. W. *J. Am. Chem. Soc.*, **1992**, *114*, 5509-5514.

7) Proudnikov, D. and Mirzabekov, A. *Nucl. Acids Res.,* **1996***, 24 (22),* 4535- 4542.

8) F. Seela, V. R. Sirivolu, P. Chittepu. *Bioconjugate Chem*. **2008**, *19*, 211

9) Meyer, K. L. and Hanna, M. M. *Bioconjugate Chem.*, **1996**, *7(4),* 401–412.

10) Platen, M. and Steckhan, E. *Liebigs Annalen der Chemie*, **1984**, *9*, 1563-1576.

11) Wallace, O. B. and Springer, D. M. *Tet. Let.* **1998**, *39(18)*, 2693-2694.

12) Nishimura, O., Kitada, C., and Fujino, M. *Chem. Pharm. Bull*. **1978**, *26(5)*, 1576.

13) Gordon, E. M., Godfrey, J. D., Delaney, N. G., Asaad, M. M., Von Langen, D., Cushman, D. W. *J. Med. Chem.* **1988,** *31(11),* 2199.

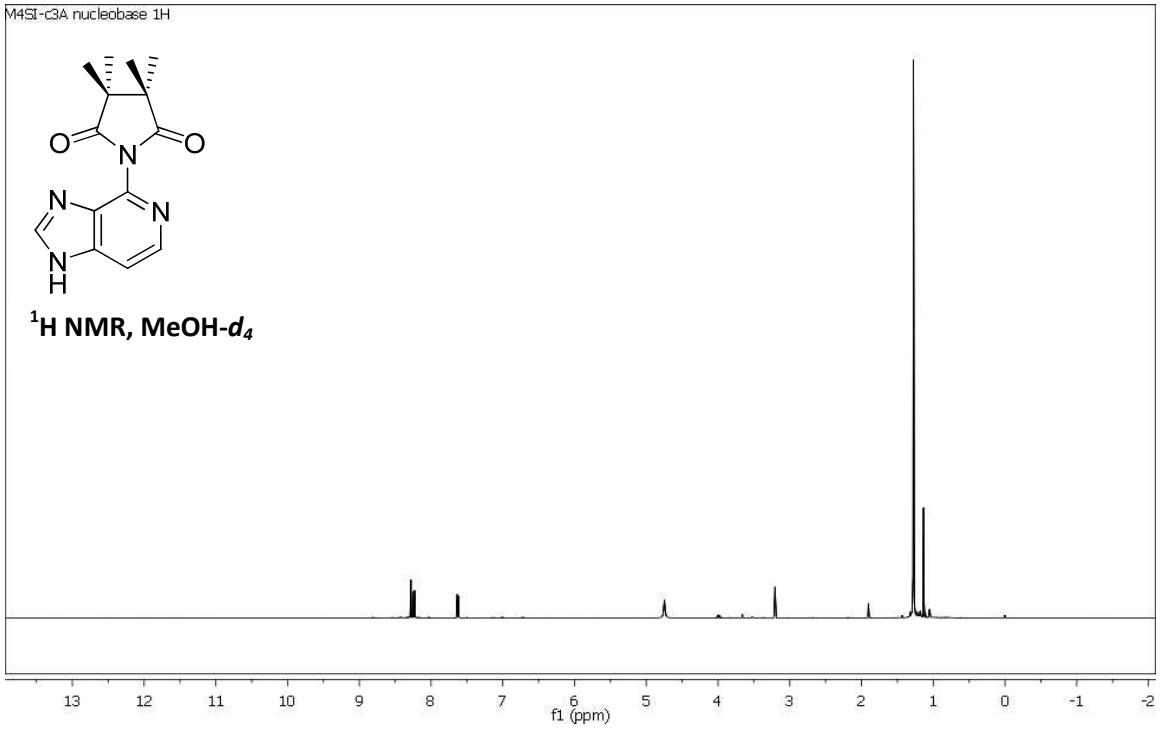
14) Christopherson, M. S. and Broom, A. D. *Nucleic Acids Res.,* **1991**, *19(20)*, 5719-5724.

15) Tran, T. K., Vuillaume, D. et al. *Chem. Eur. J.,* **2008**, *14*, 6237-6246.

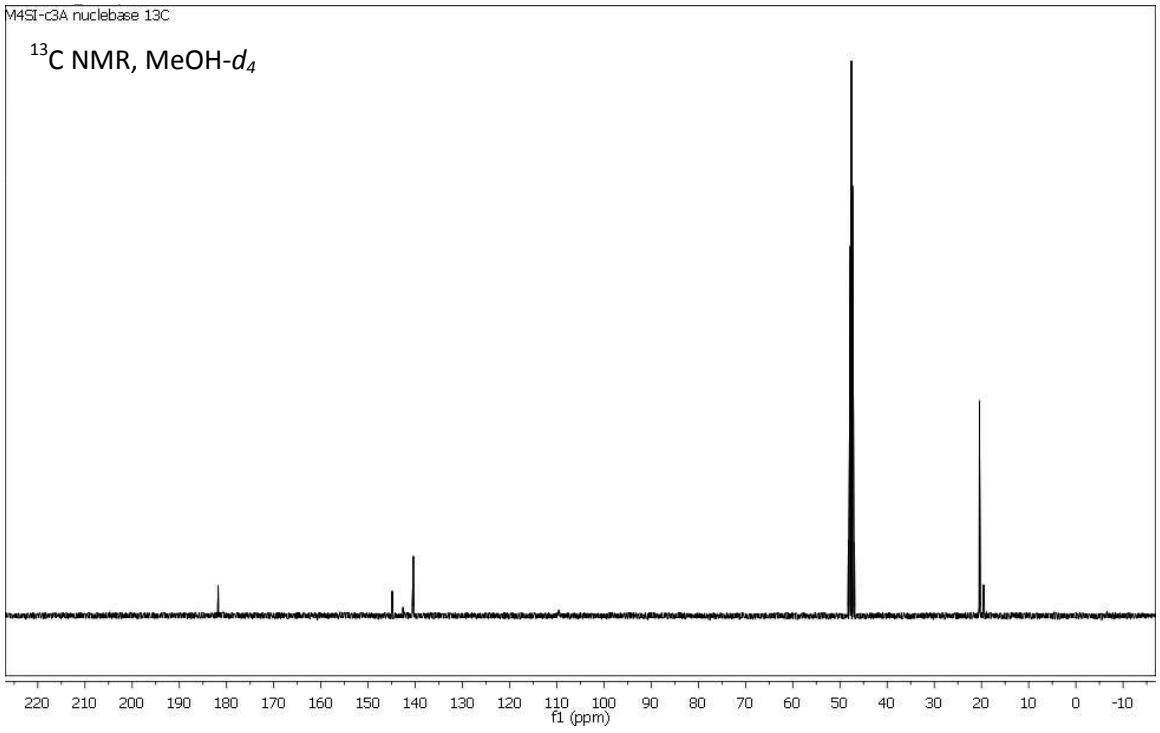
16) Puigmarti-Luis, J., Amabilino, D. B., et al. *Angew. Chem. Int. Ed.* **2008**, *47*, 1861 –1865.

17) Clivio, P., Woissard, A. et al. *Tet. Let.,* **1992***, 33(1)*, 65-68.

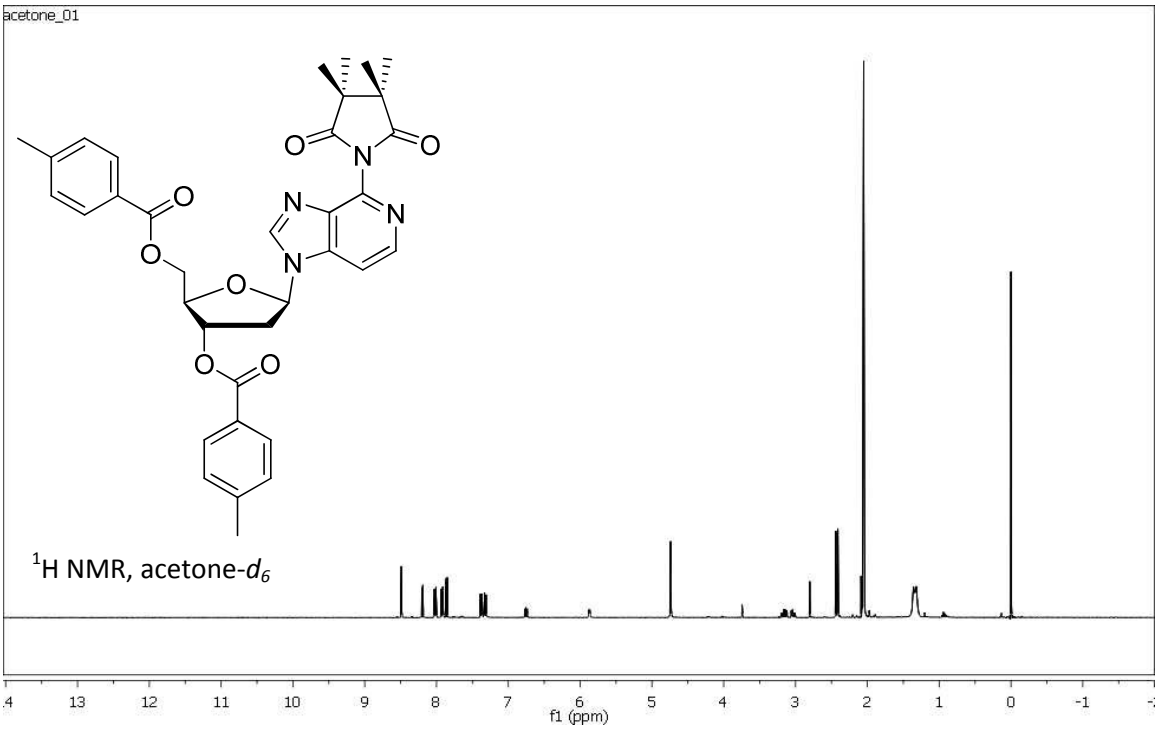
Appendix 1: NMR Spectra

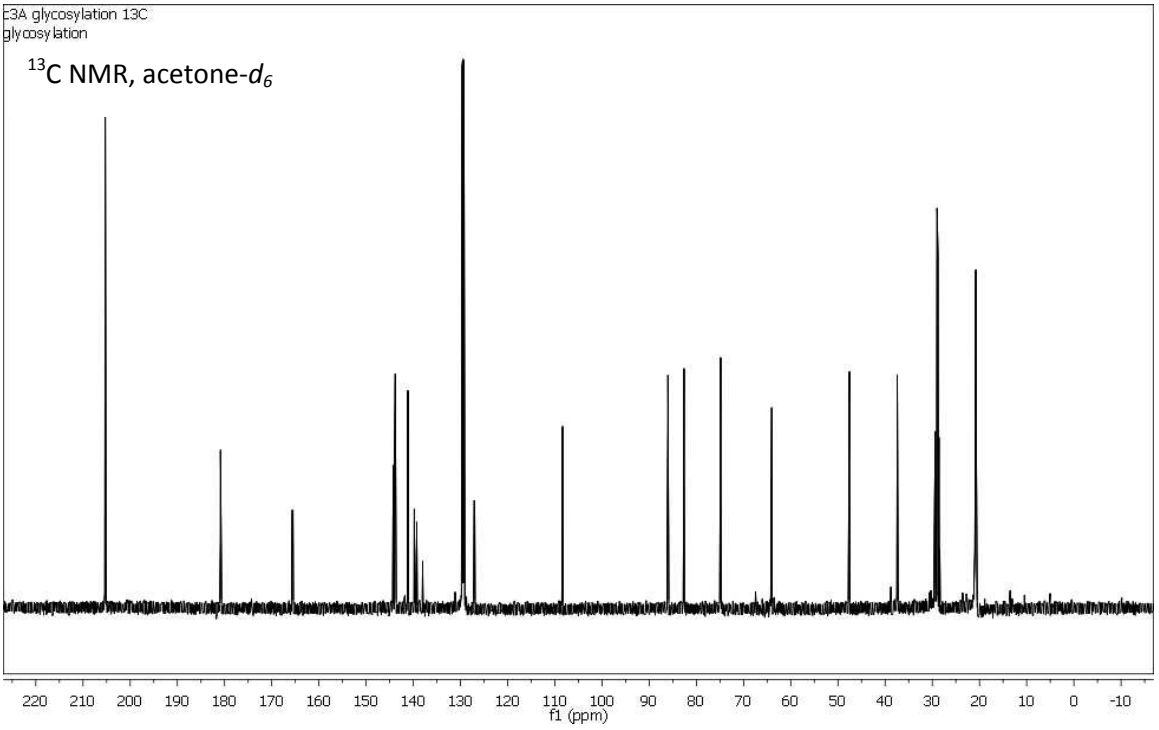


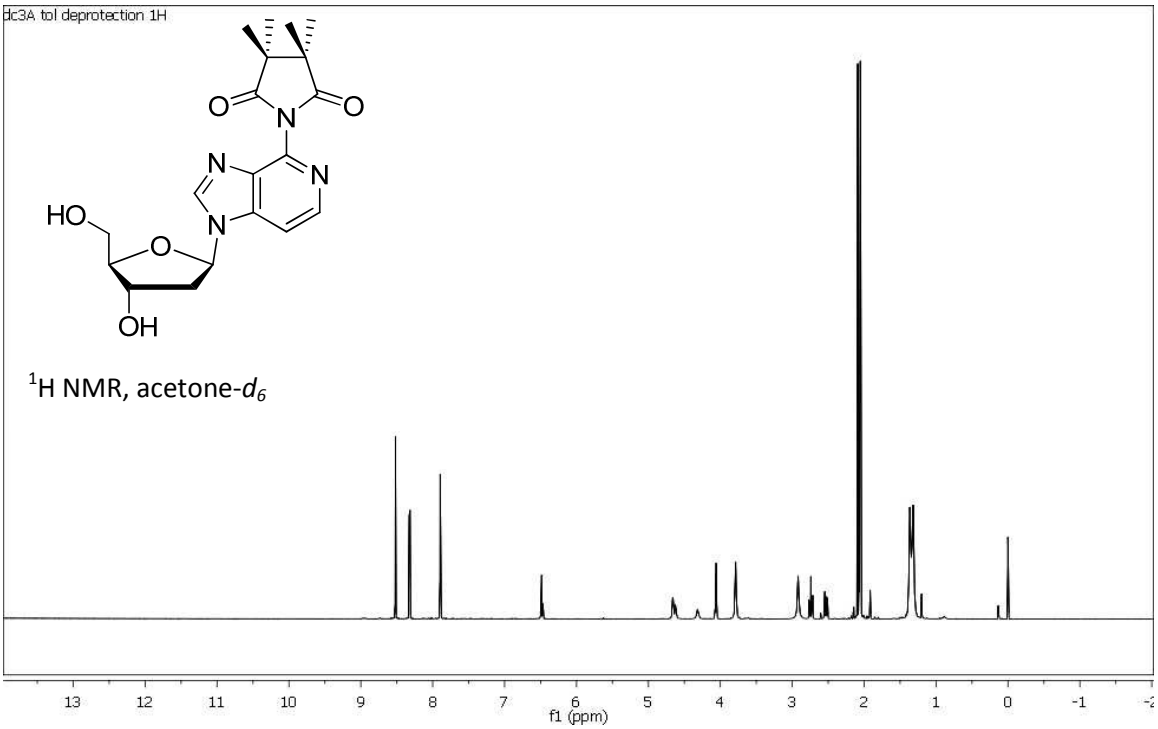
^{1}H NMR, MeOH-d_4

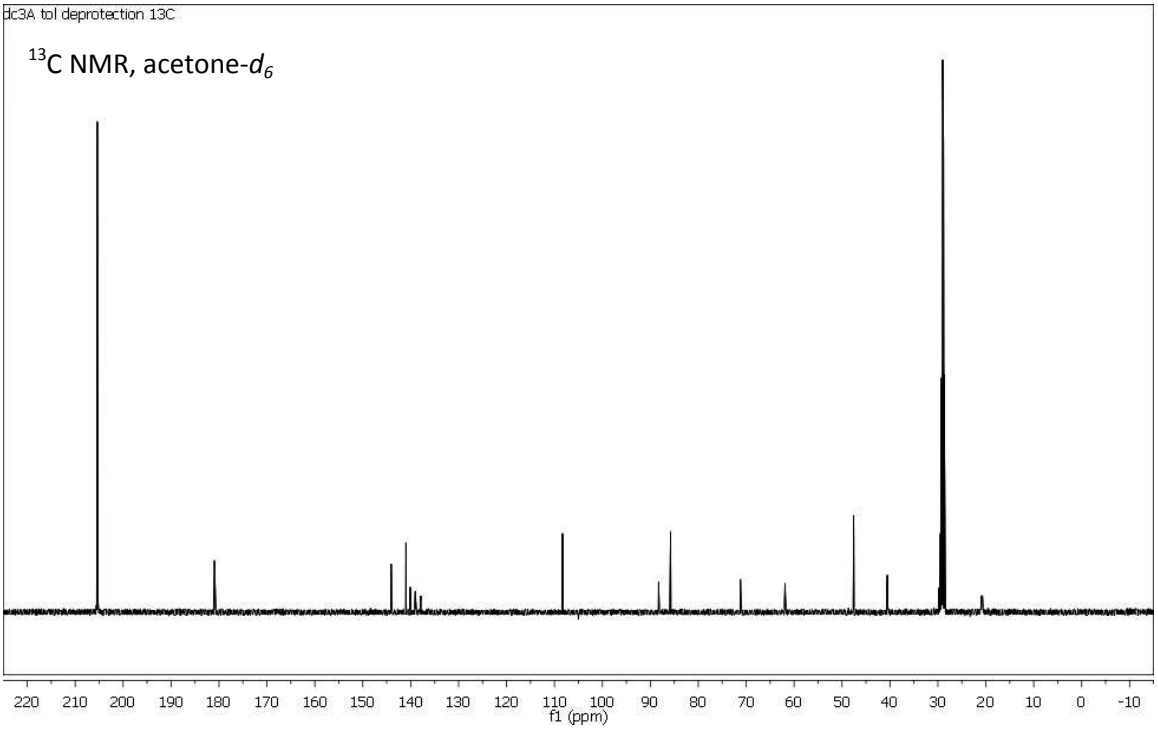


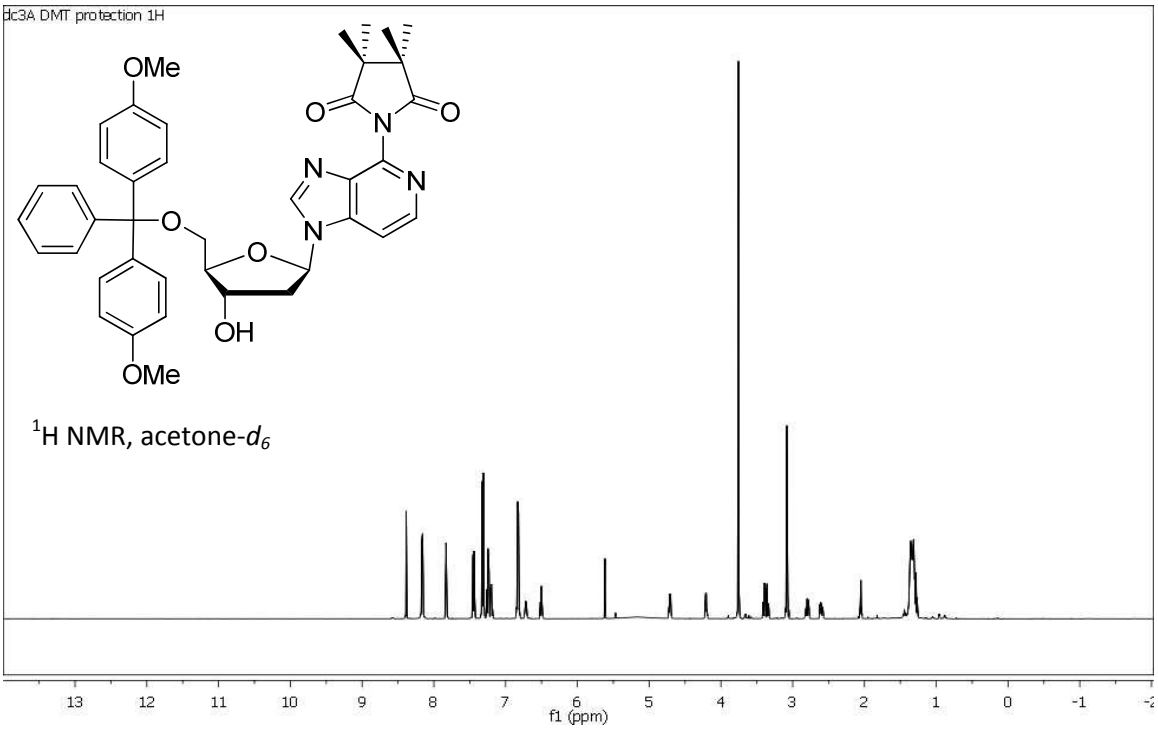
^{13}C NMR, MeOH-d_4


acetone_01
^{1}H NMR, acetone-d_6
f1 (ppm)

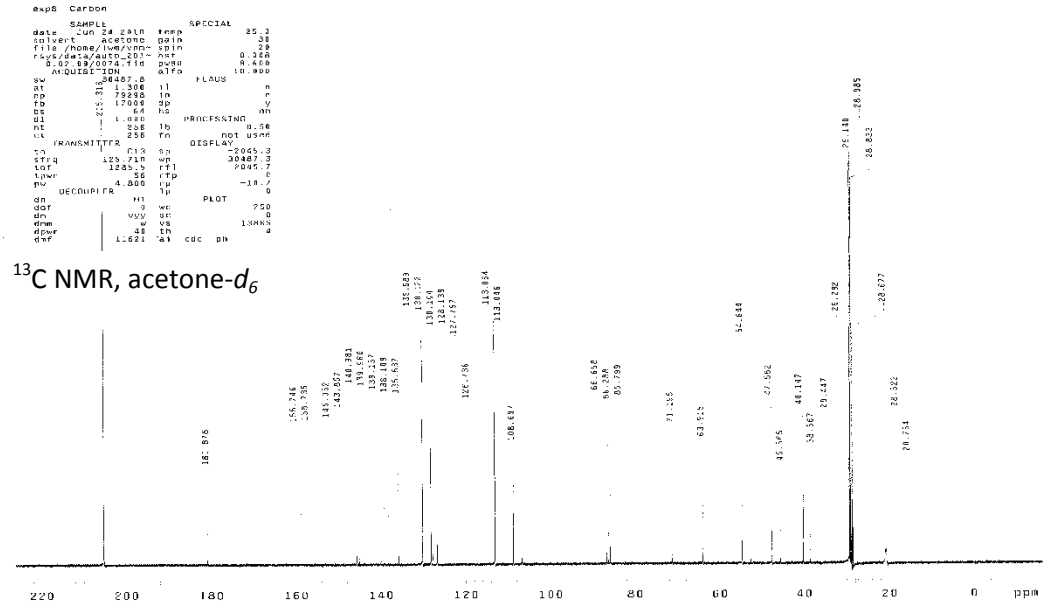

C3A glycosylation 13C
glycosylation
^{13}C NMR, acetone-d_6
f1 (ppm)


dc3A tol deprotection 1H
HO
O
OH
N
N
N
N
O
O
^{1}H NMR, acetone-d_6
13 12 11 10 9 8 7 6 5 4 3 2 1 0 -1
f1 (ppm)

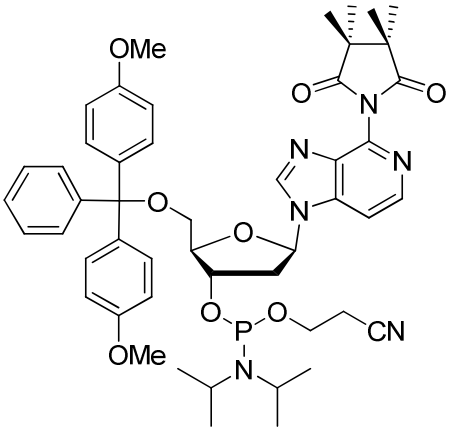

dc3A tol deprotection 13C
^{13}C NMR, acetone-d_6
220 210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
f1 (ppm)



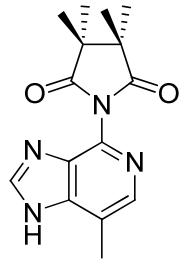
^{1}H NMR, acetone-d_6



^{13}C NMR, acetone-d_6

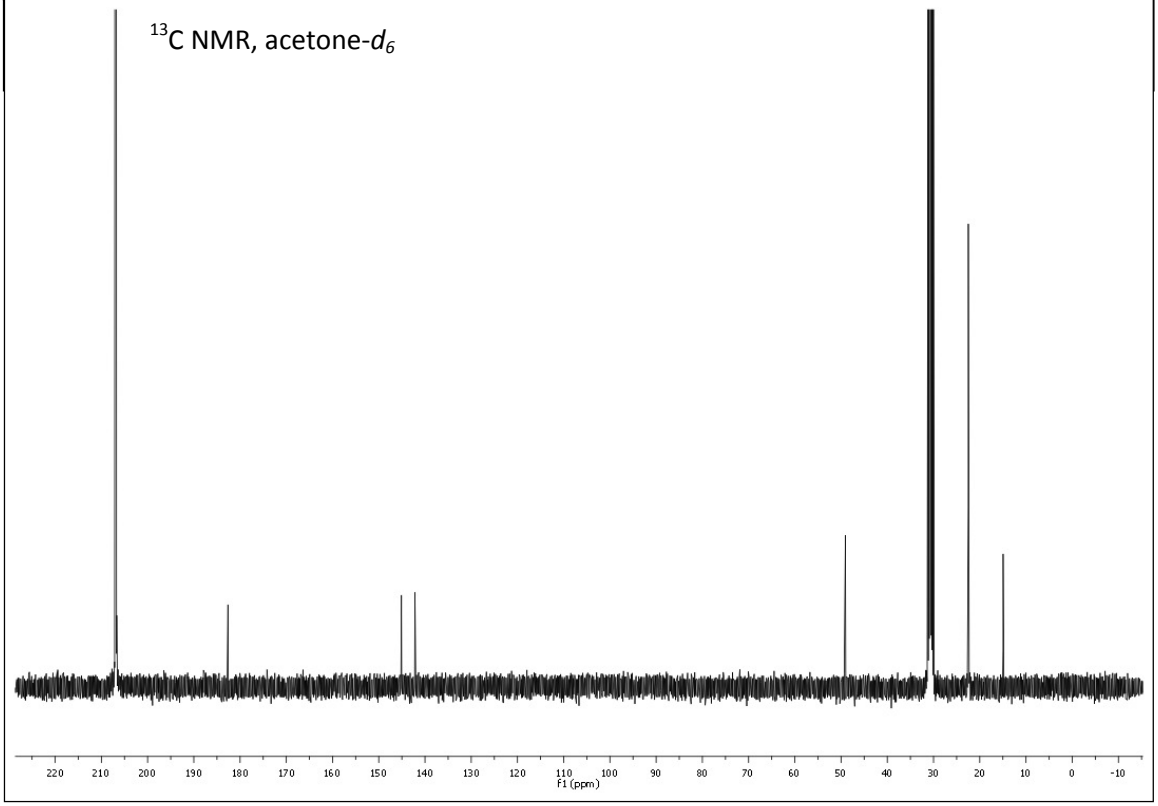


^{31}P NMR, acetone-d_6

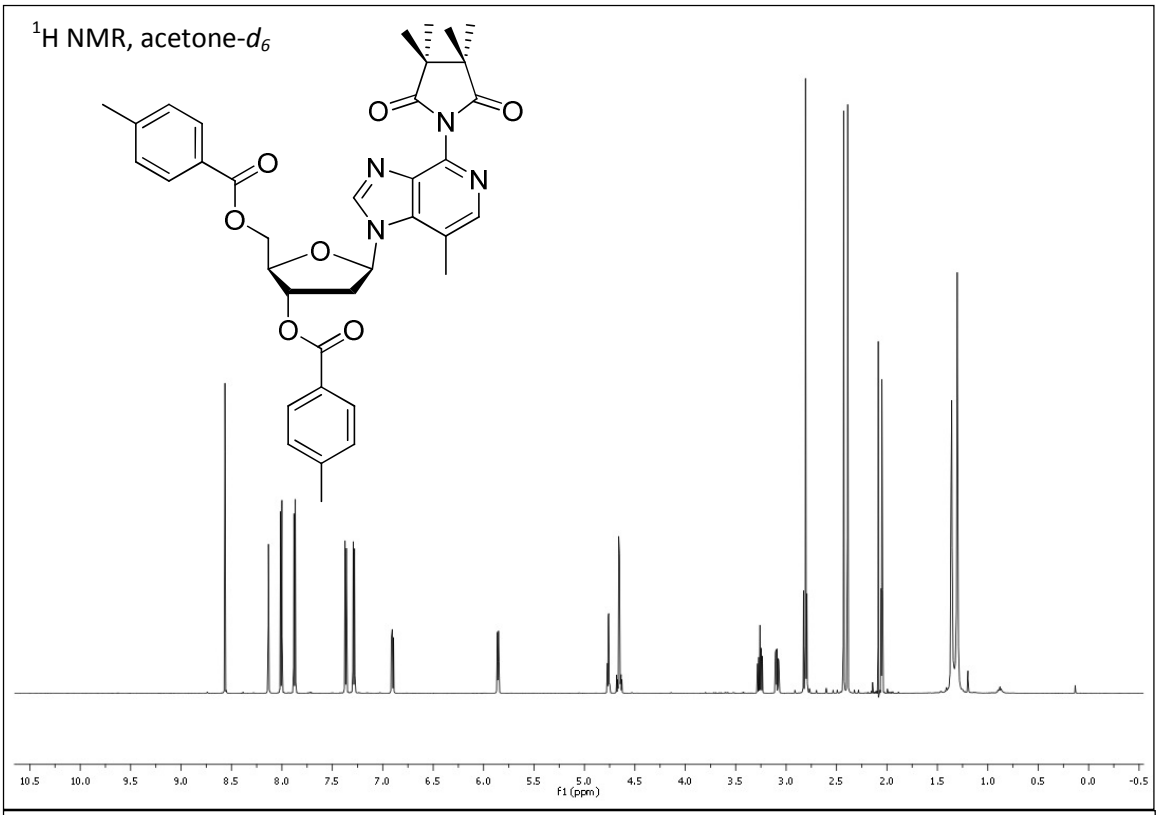


^{1}H NMR, acetone-d_6

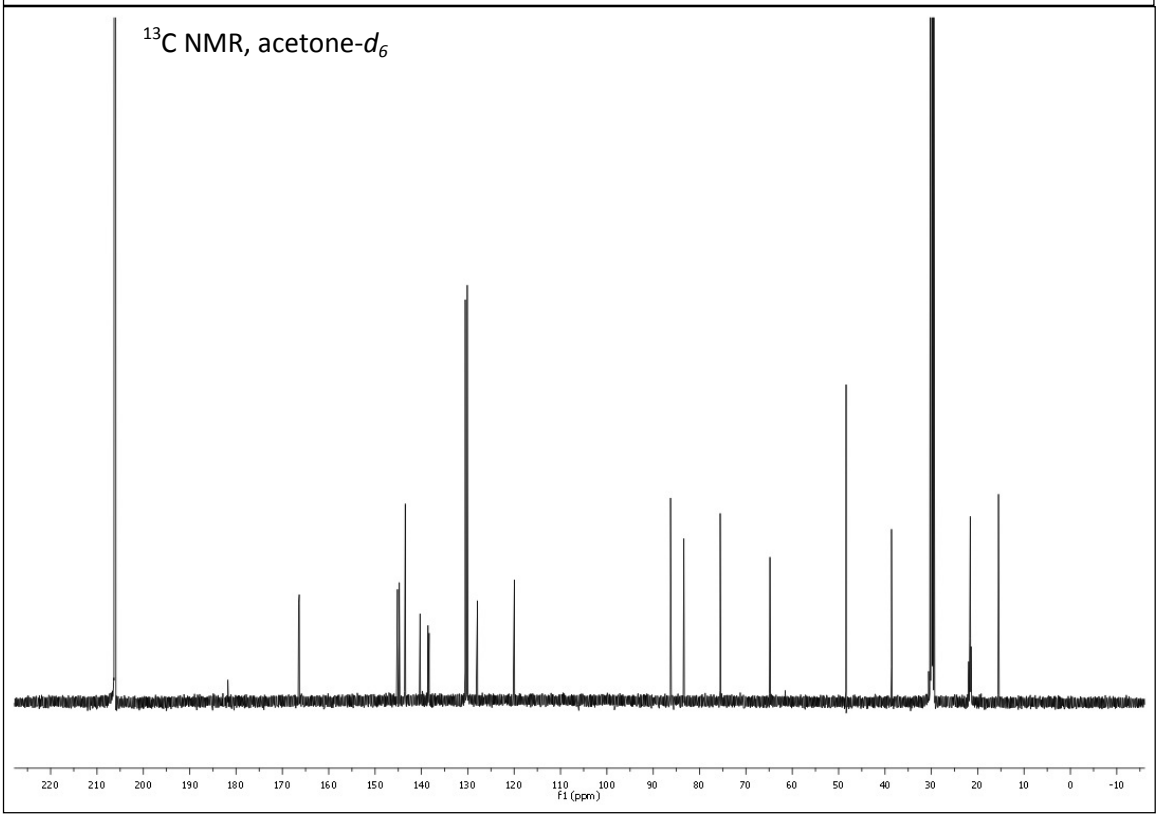
^{13}C NMR, acetone-d_6

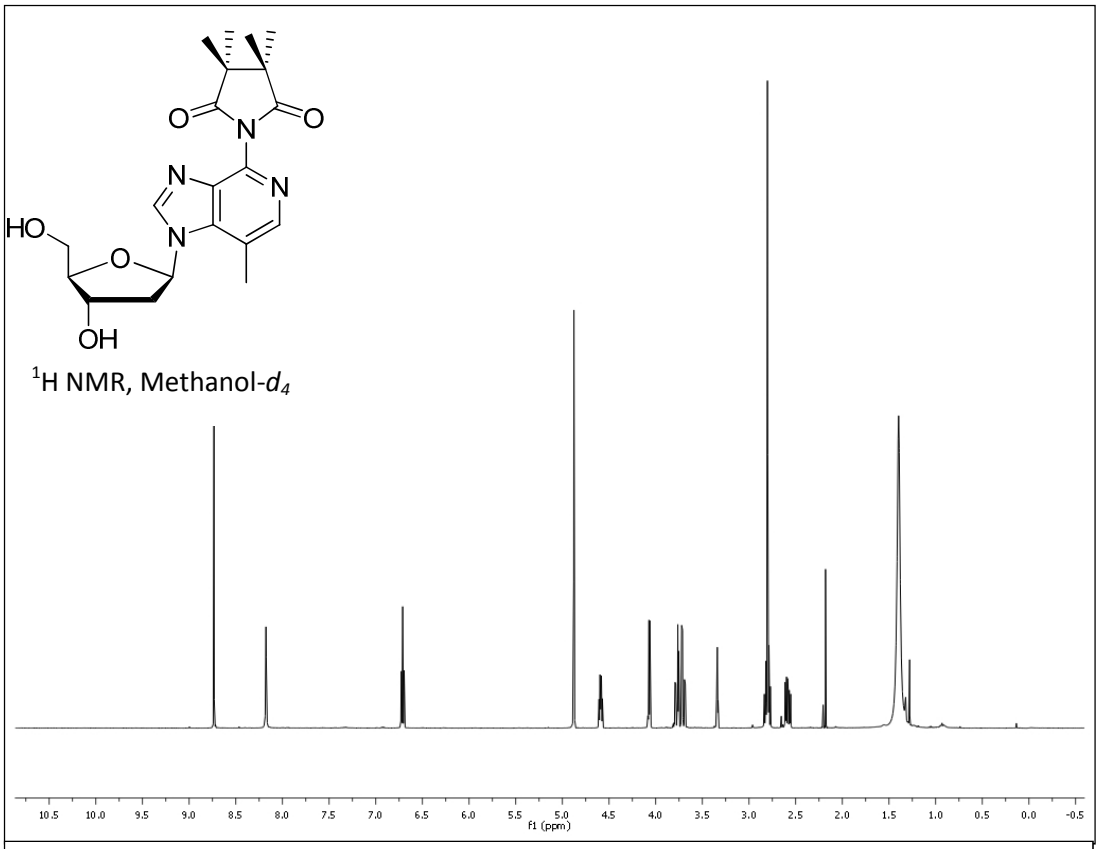


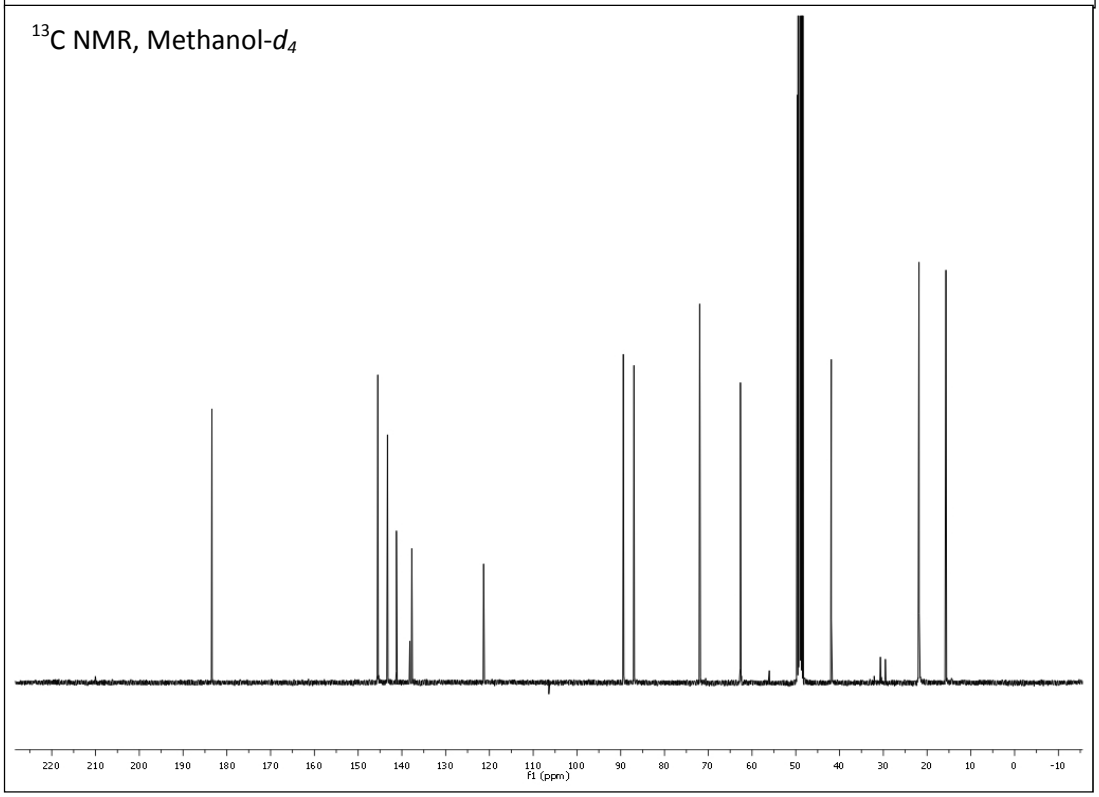
^{1}H NMR, acetone-d_6

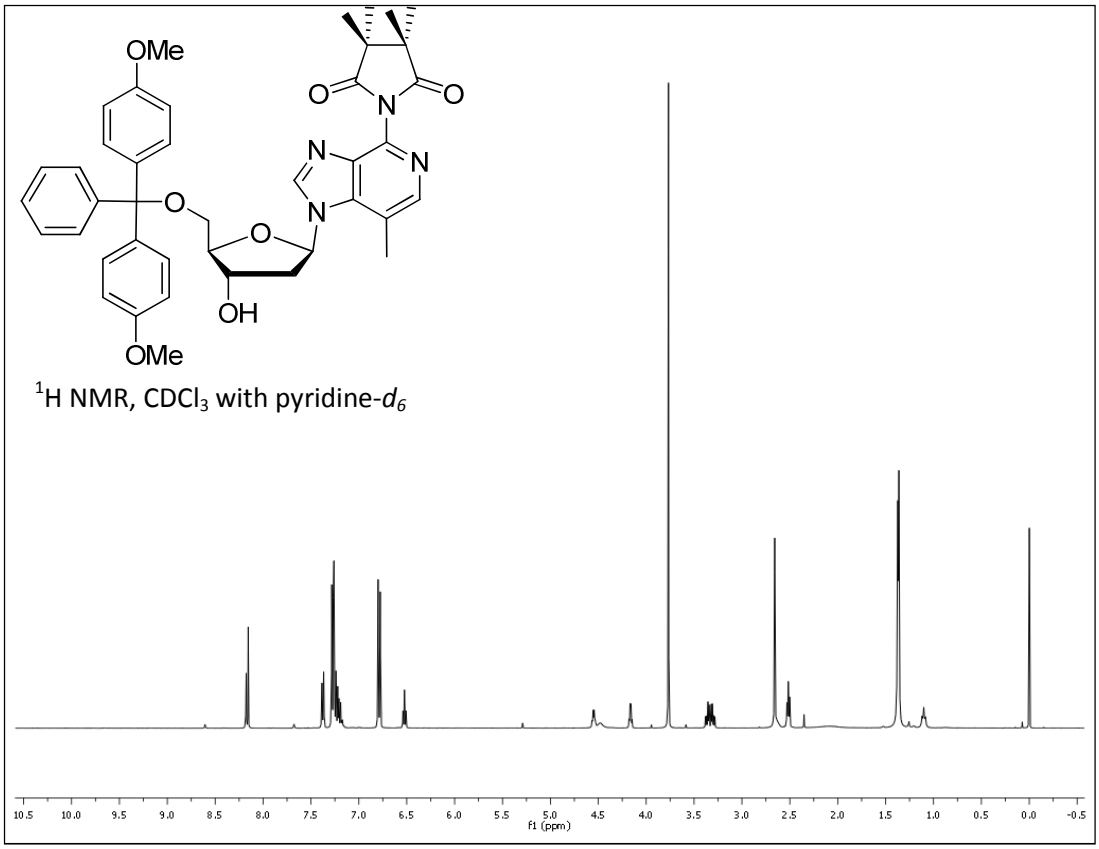


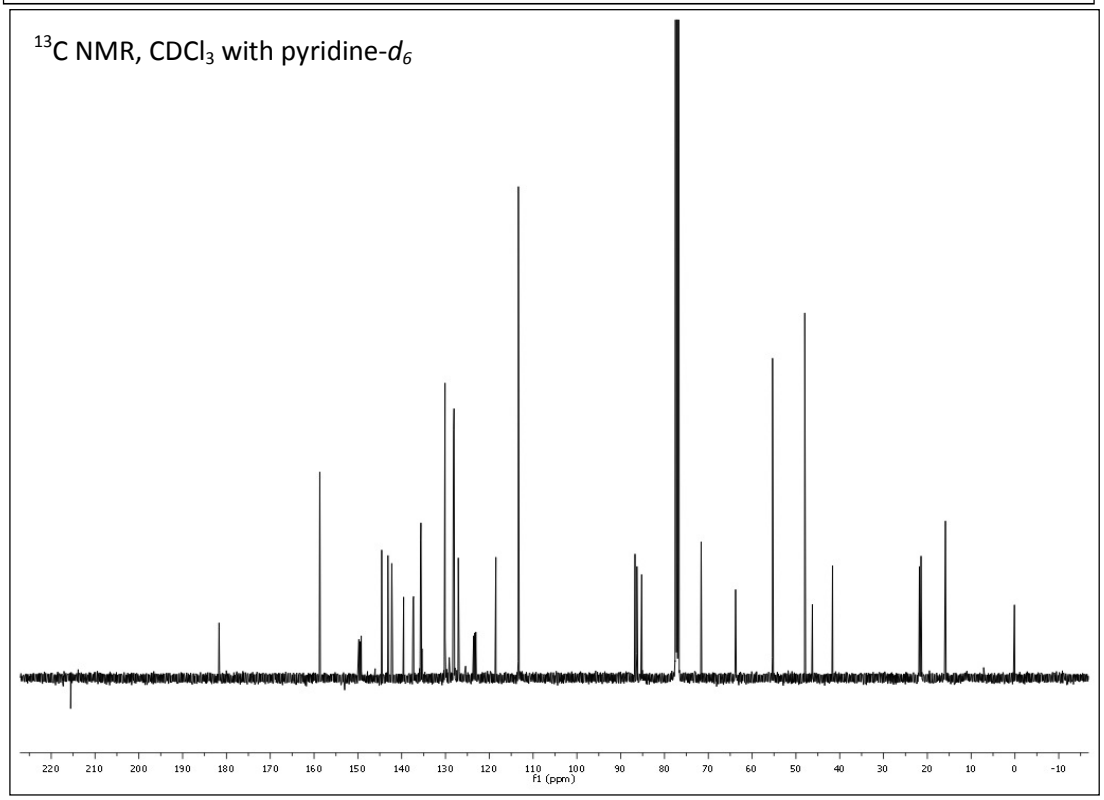
^{13}C NMR, acetone-d_6

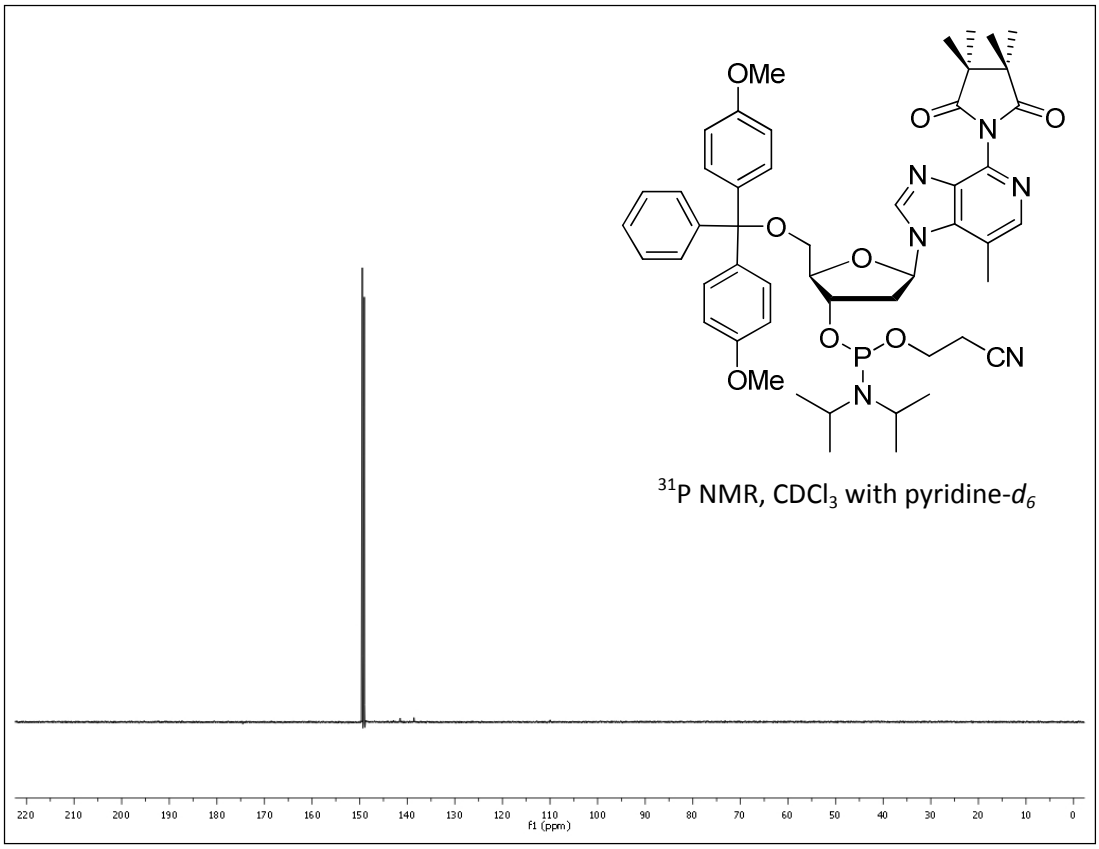



HO
O
N
O
O
N
N
N
OH
^{1}H NMR, Methanol-d_4
10.5 10.0 9.5 9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 0.0 -0.5
f1 (ppm)

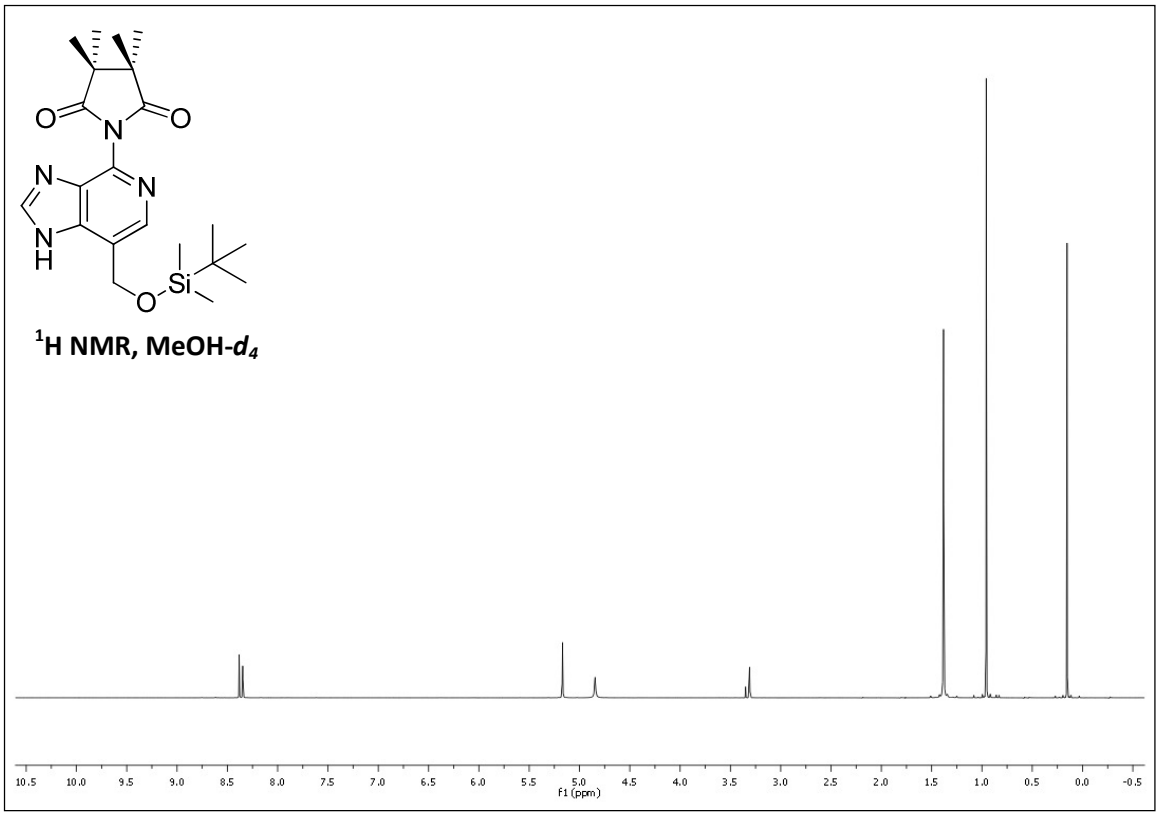

^{13}C NMR, Methanol-d_4
220 210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
f1 (ppm)

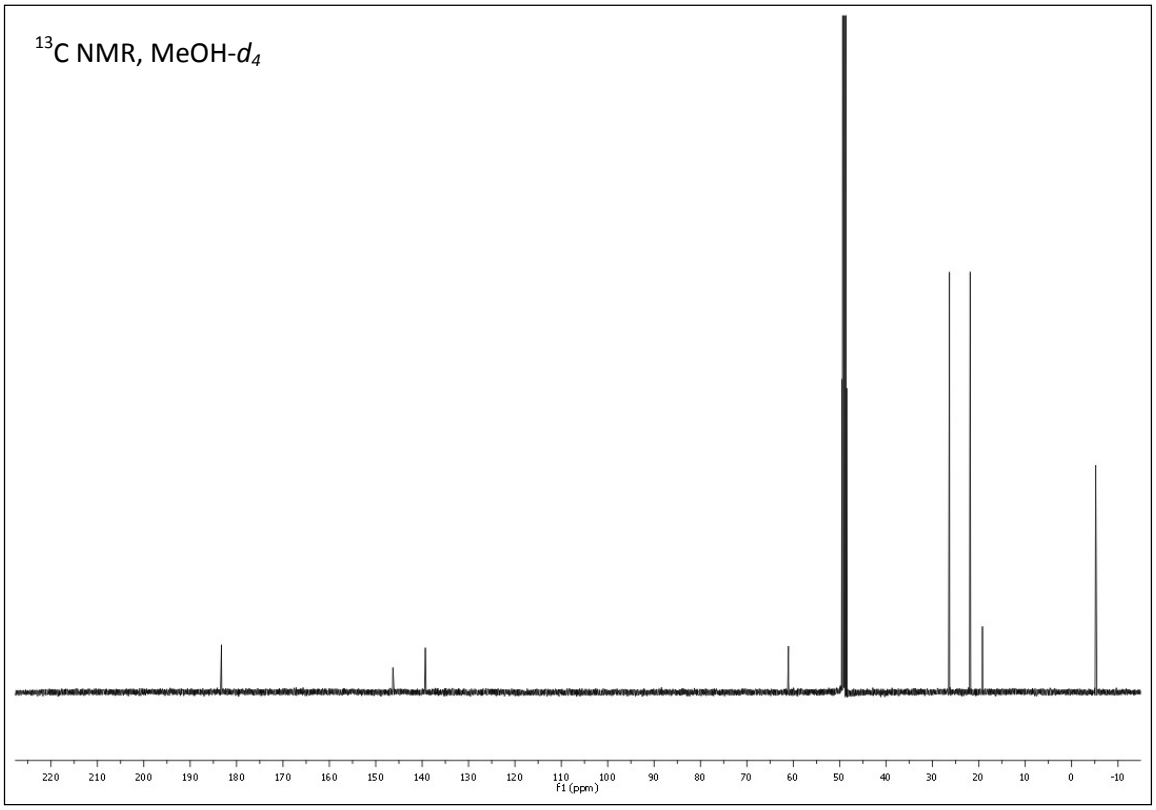

OMe
O
N
O
N
N
N
O
O
N
OH
OMe
^{1}H NMR, $CDCl_3$ with pyridine-d_6
10.5 10.0 9.5 9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 0.0 -0.5
f1 (ppm)

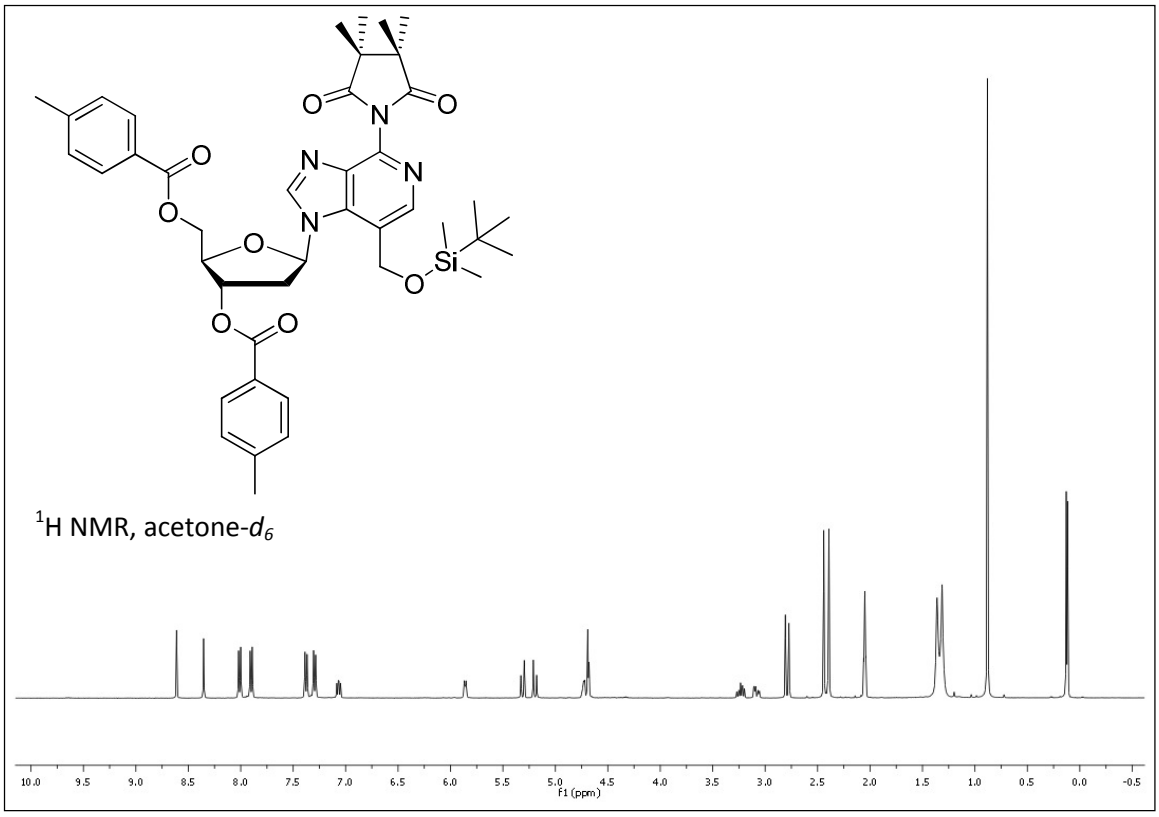

^{13}C NMR, $CDCl_3$ with pyridine-d_6
220 210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
f1 (ppm)

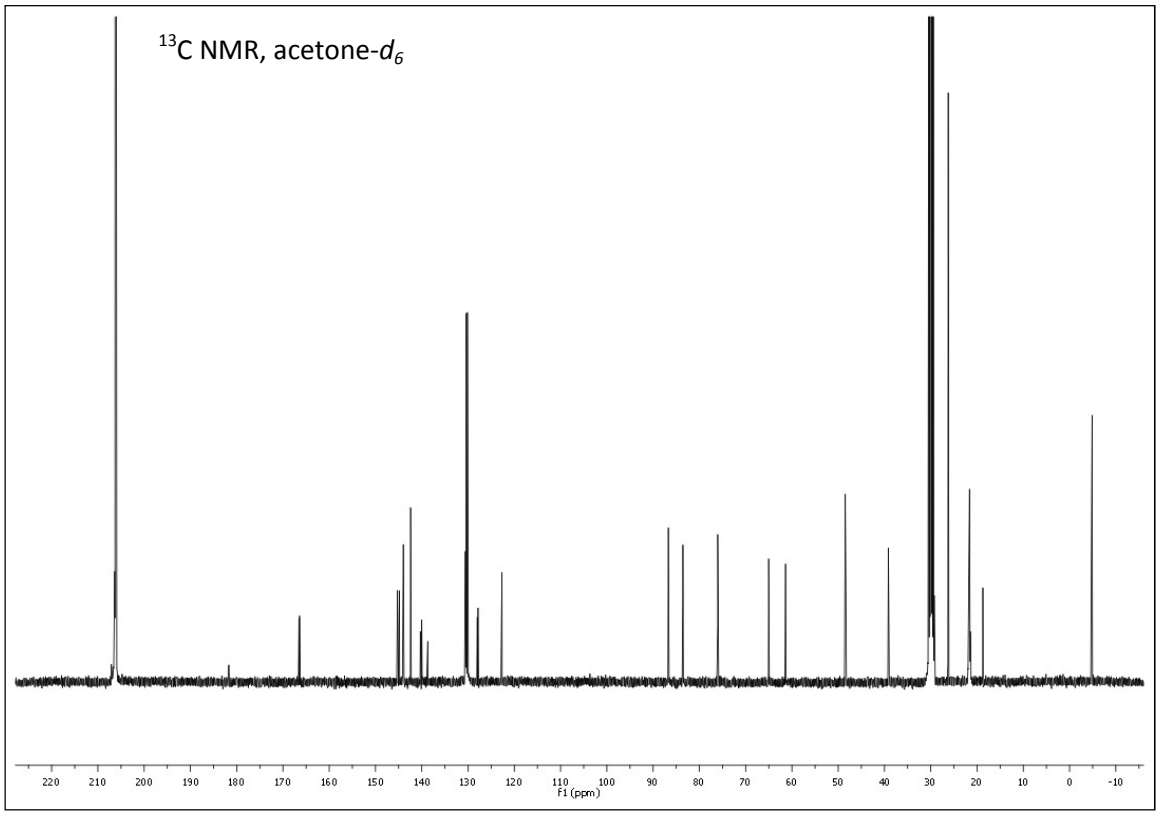


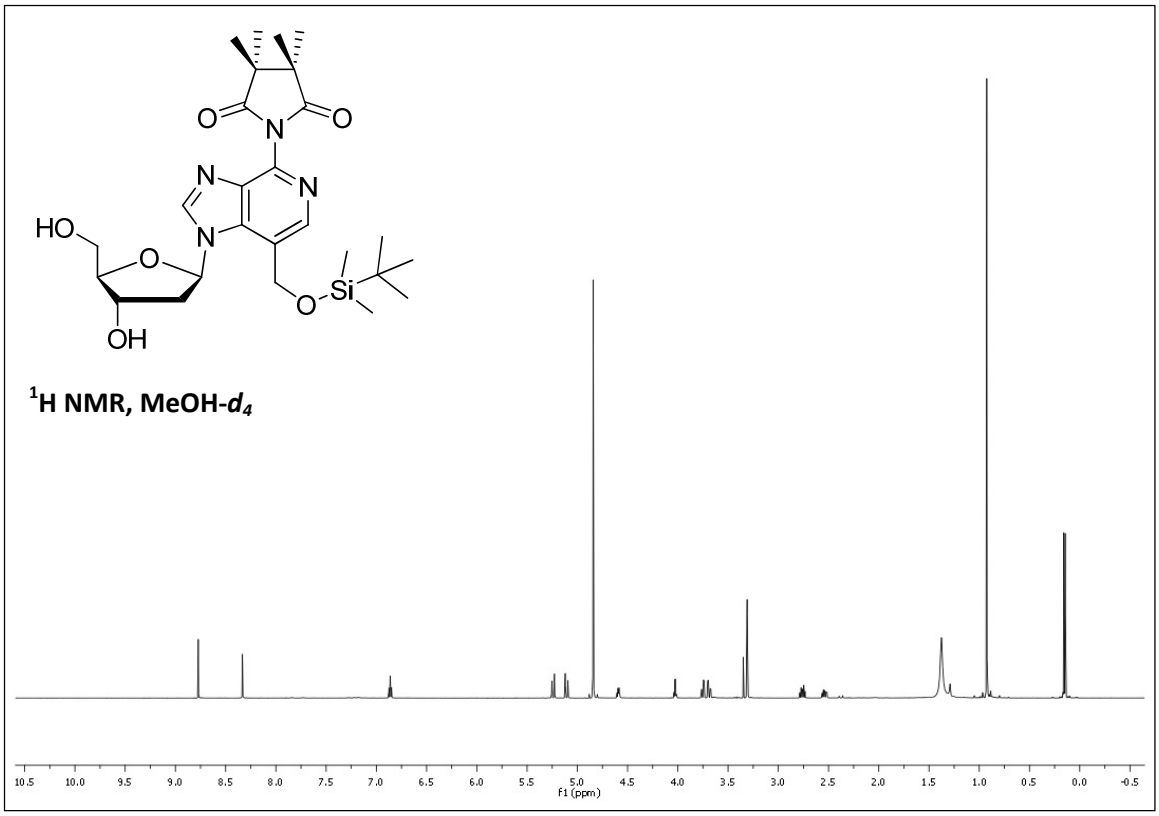
^{31}P NMR, $CDCl_3$ with pyridine-d_6

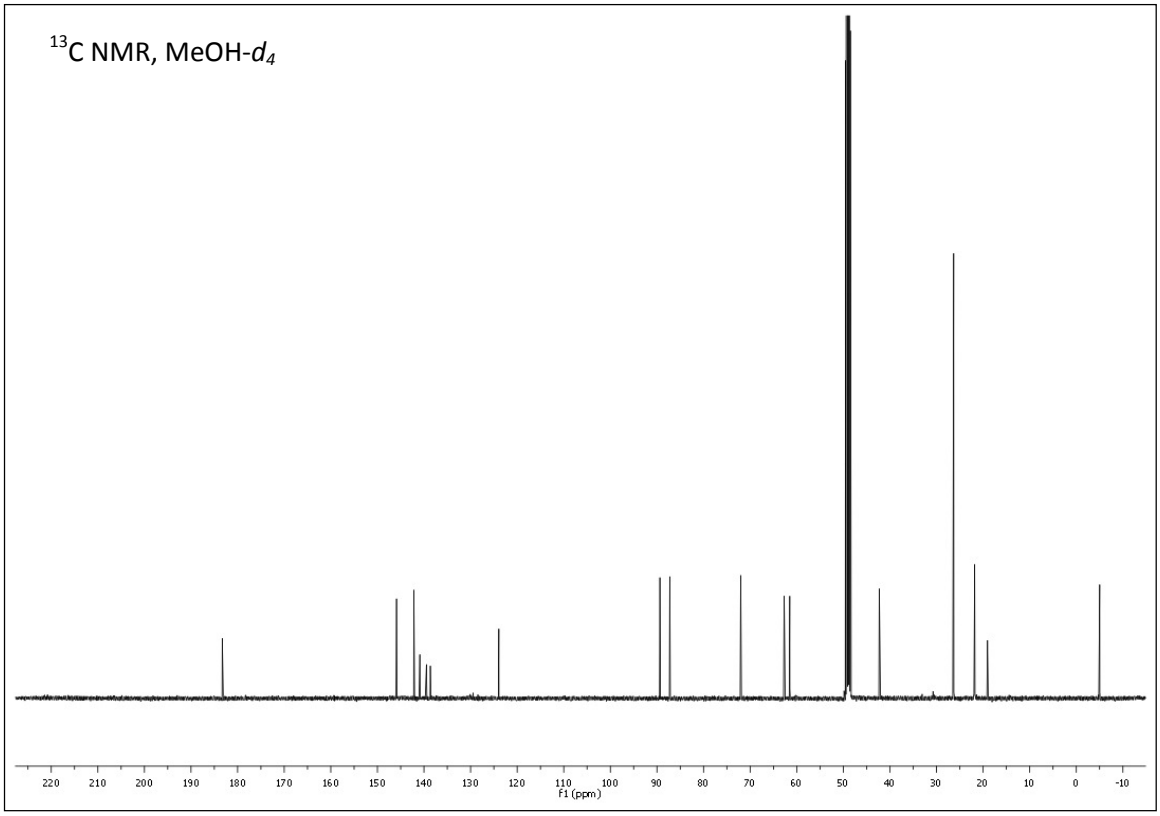

O
N
O
N
N
H
N
Si
O
^{1}H NMR, MeOH-d_4
10.5 10.0 9.5 9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 0.0 -0.5
f1 (ppm)

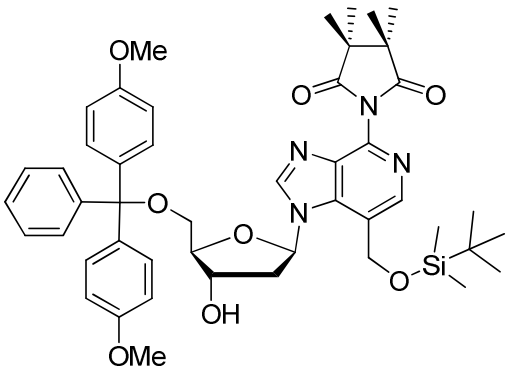

^{13}C NMR, MeOH-d_4
220 210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
f1 (ppm)


^{1}H NMR, acetone-d_6
F1 (ppm)


^{13}C NMR, acetone-d_6
F1 (ppm)

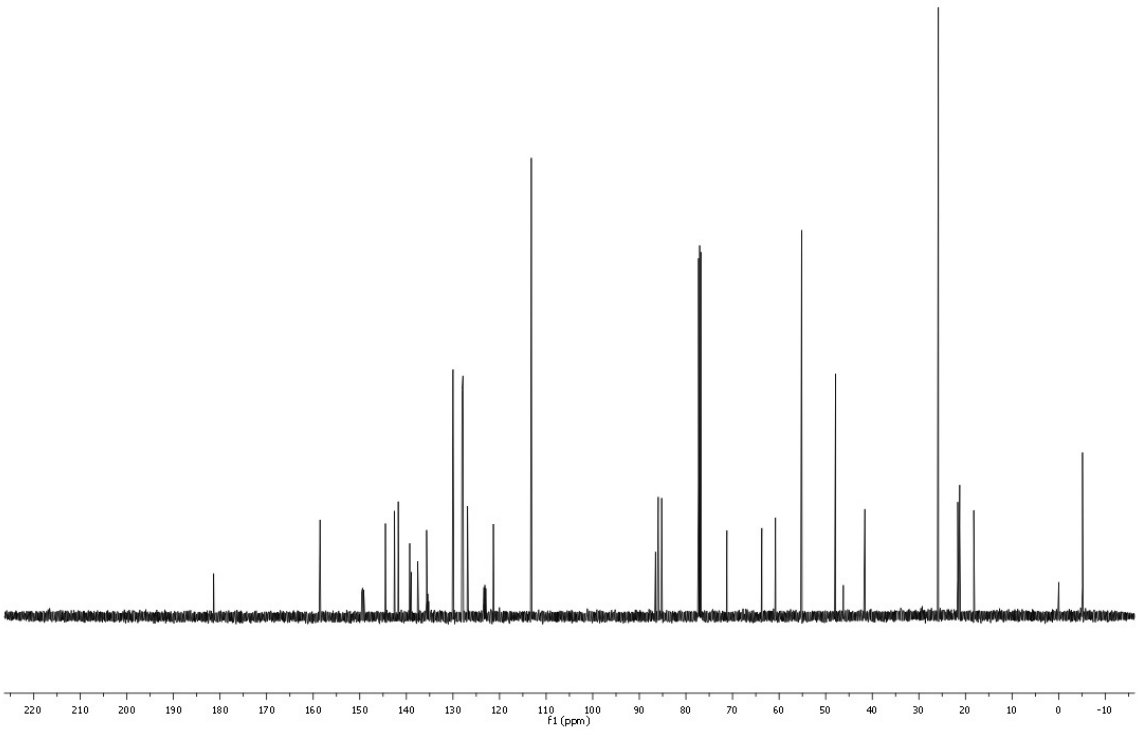

HO
O
N
O
N
N
N
O
Si
OH
^{1}H NMR, MeOH-d_4
10.5 10.0 9.5 9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 0.0 -0.5
f1 (ppm)

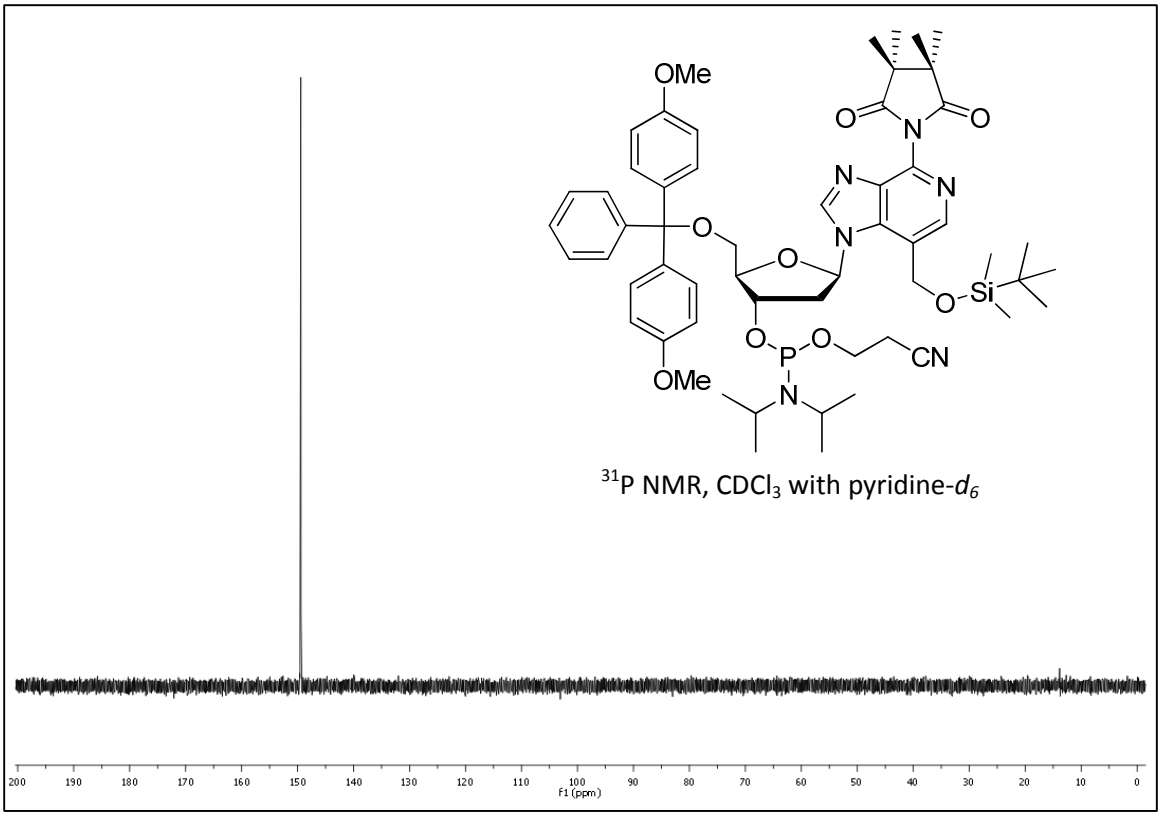

^{13}C NMR, MeOH-d_4
220 210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
f1 (ppm)



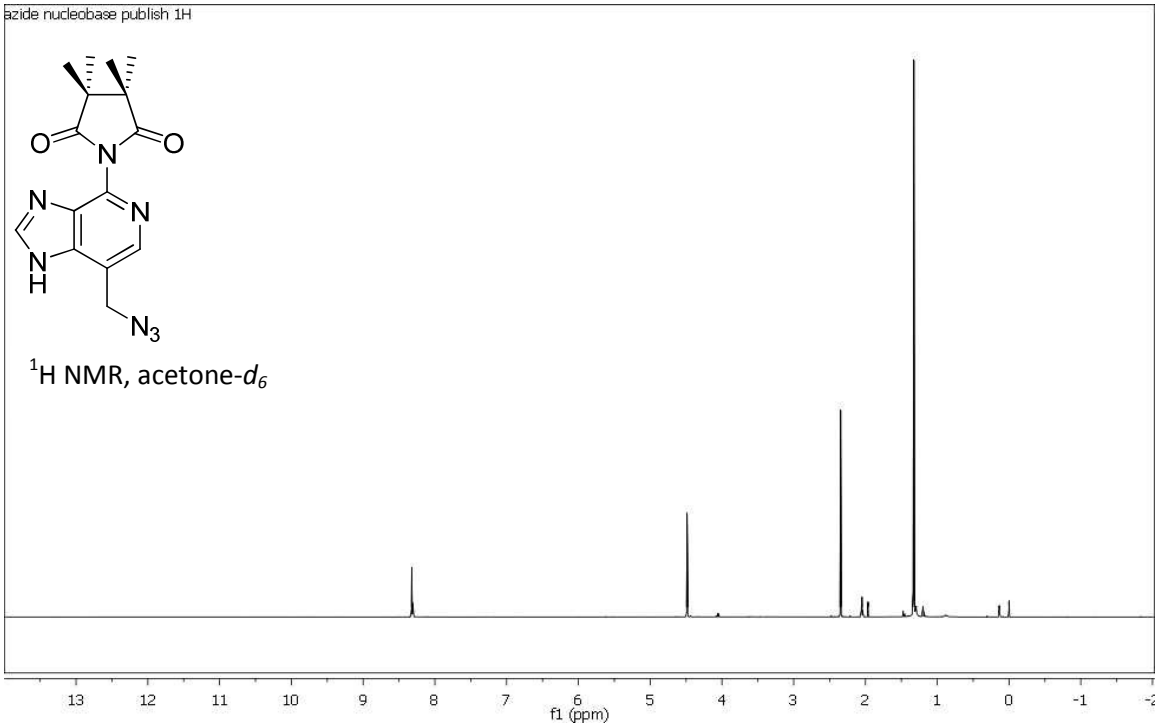
^{1}H NMR, $CDCl_3$ with pyridine-d_6

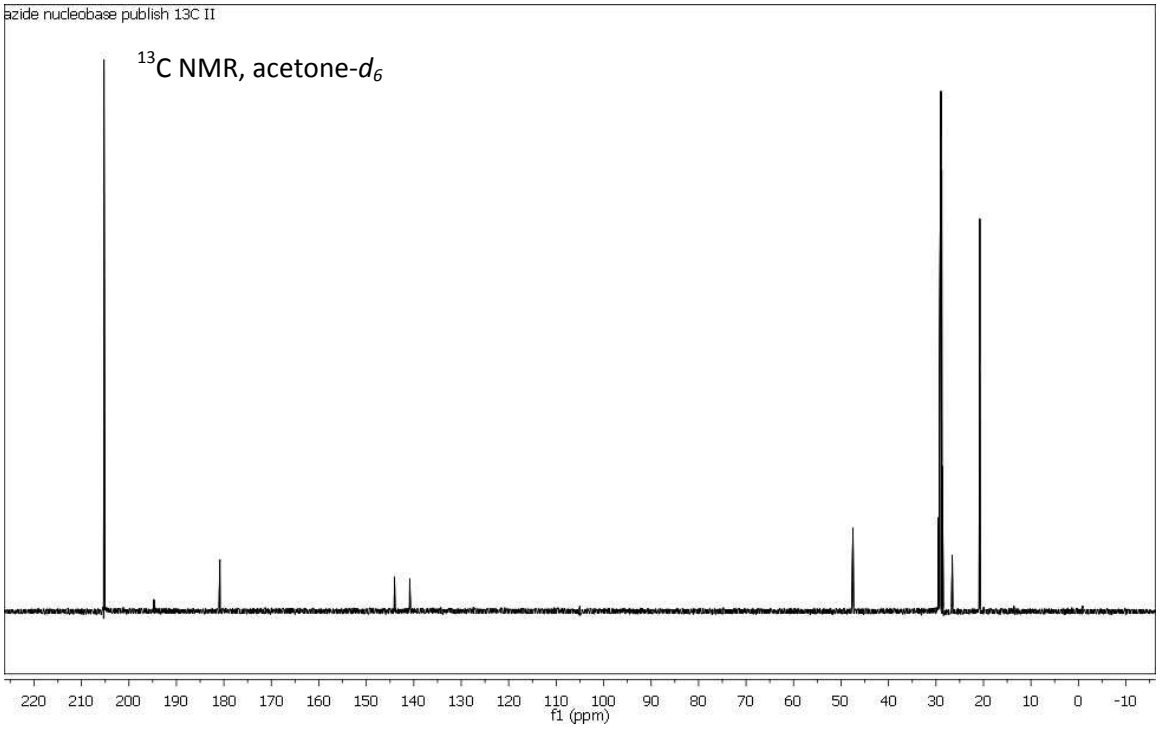
^{13}C NMR, $CDCl_3$ with pyridine-d_6

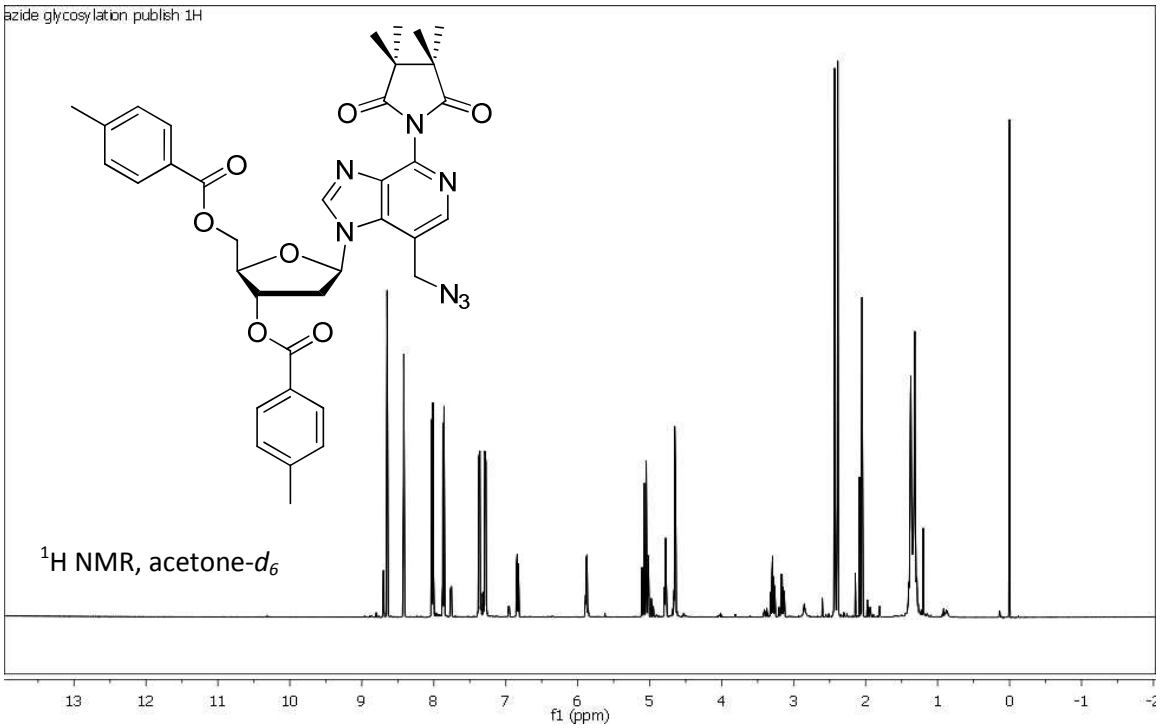


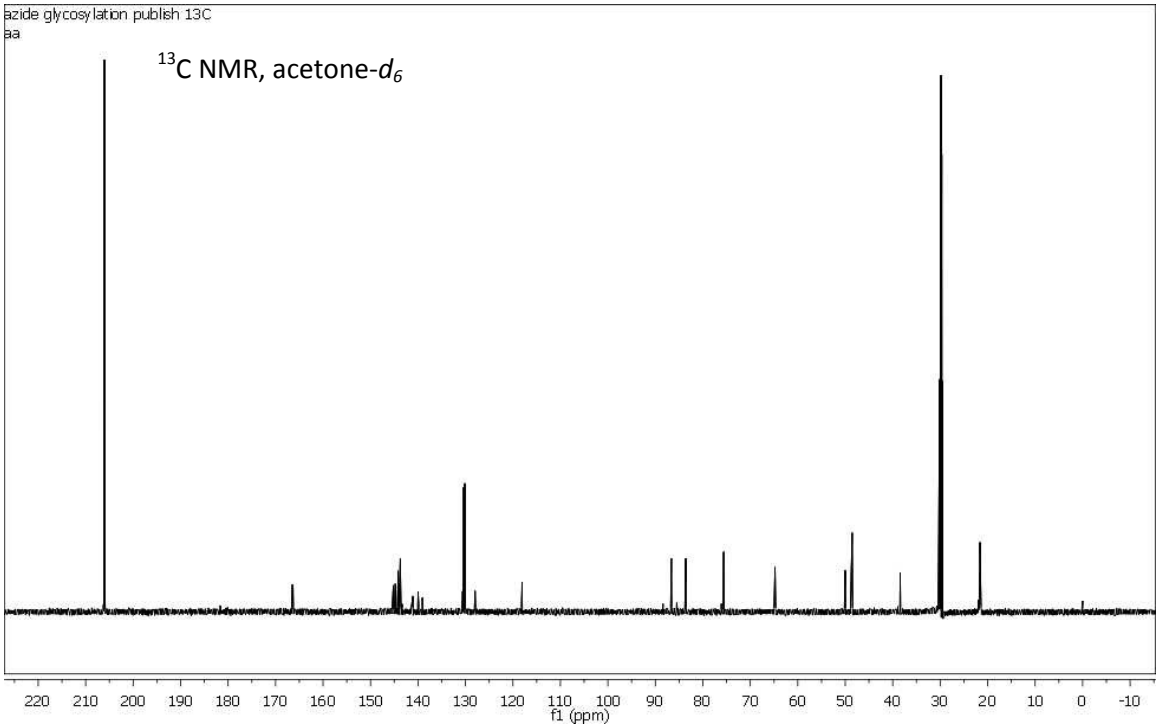


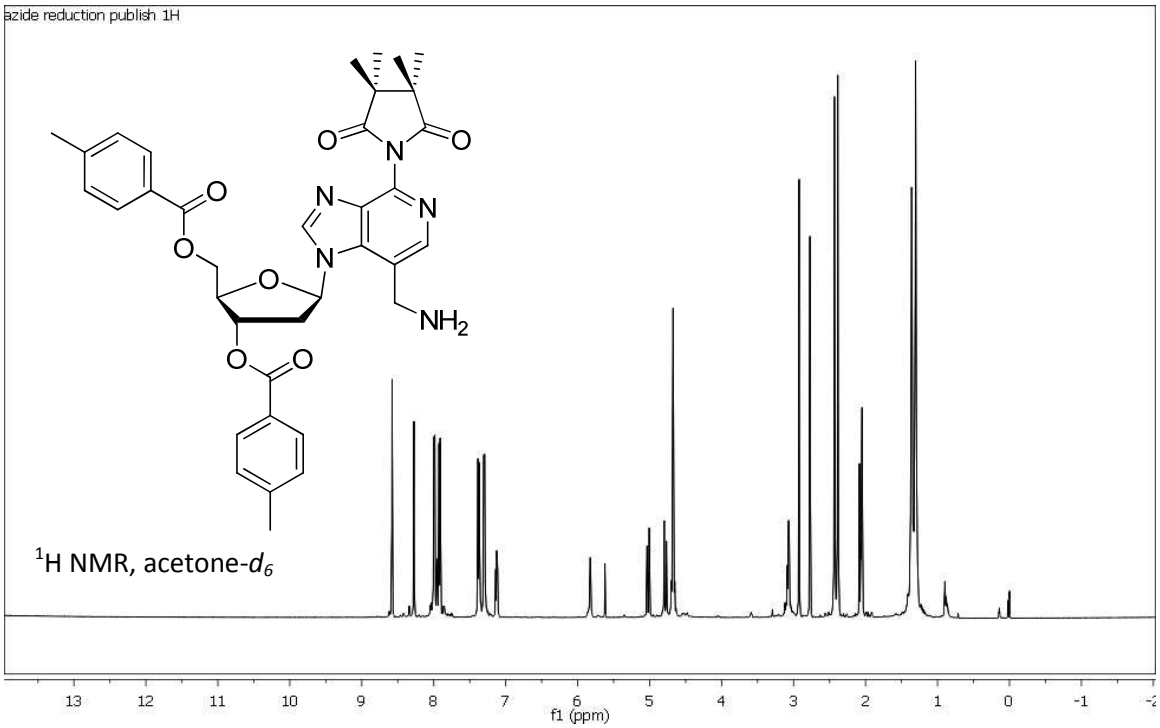
^{31}P NMR, $CDCl_3$ with pyridine-d_6

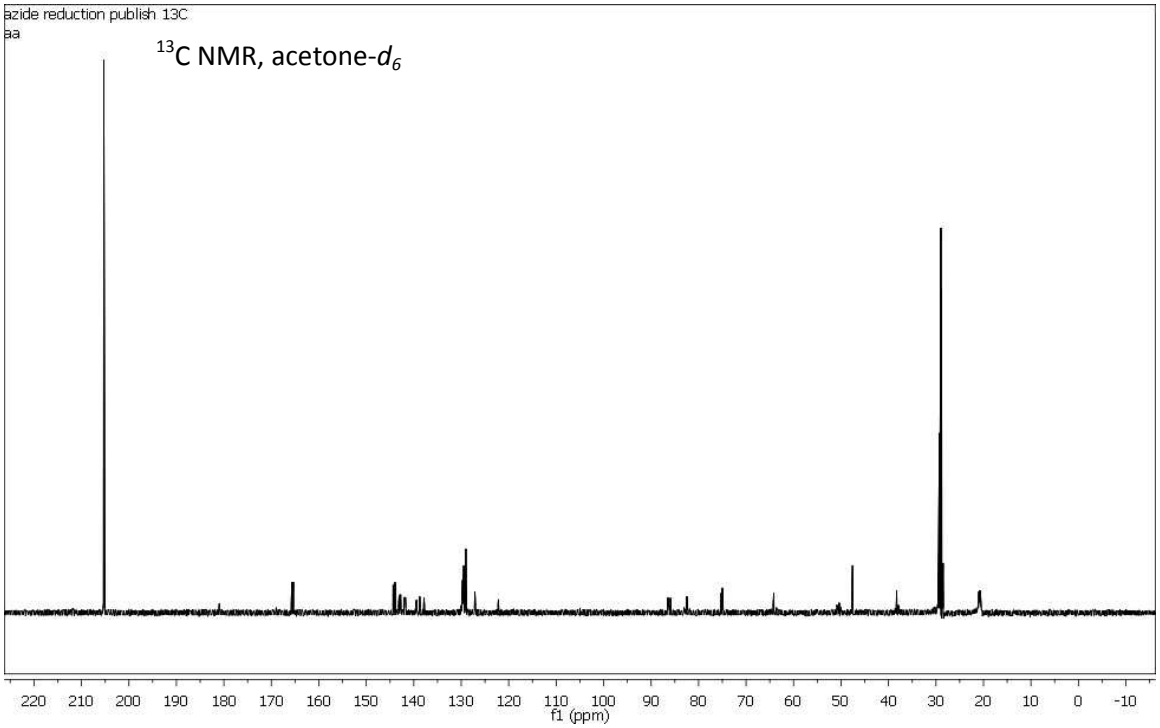

azide nucleobase publish 1H
O
N
O
N
N
H
N3
1H NMR, acetone-d6
13 12 11 10 9 8 7 6 5 4 3 2 1 0 -1
f1 (ppm)

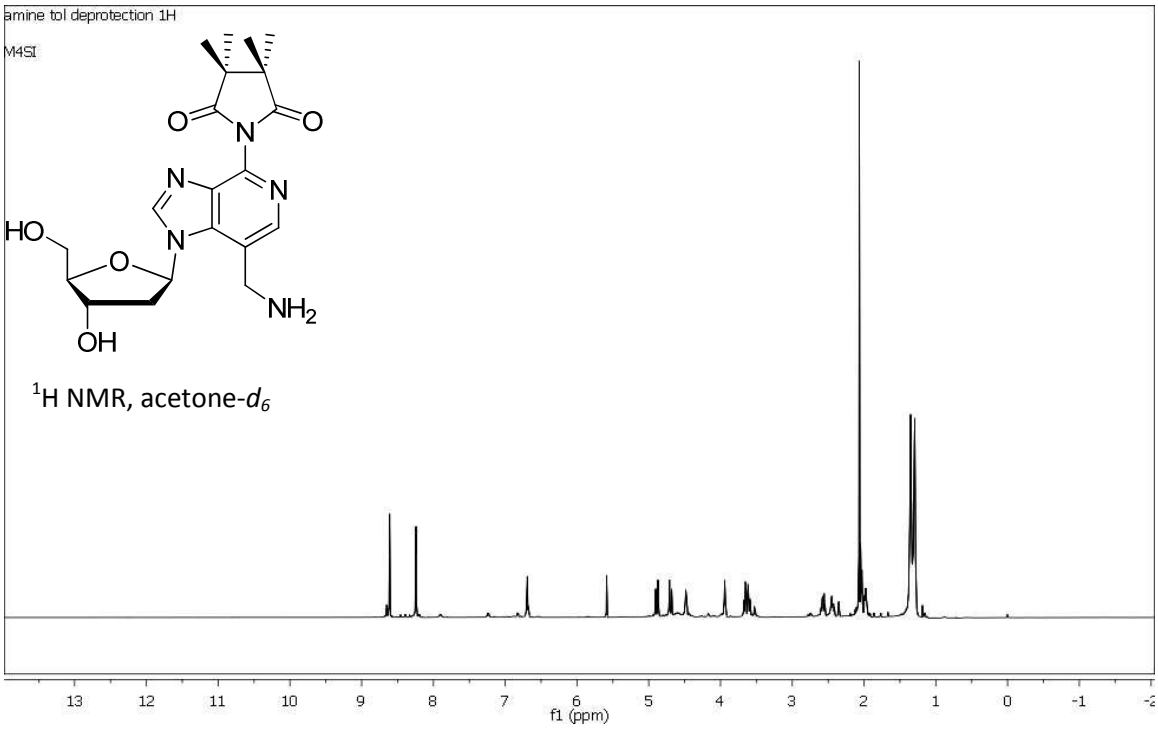

azide nucleobase publish 13C II
13C NMR, acetone-d6
220 210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
f1 (ppm)

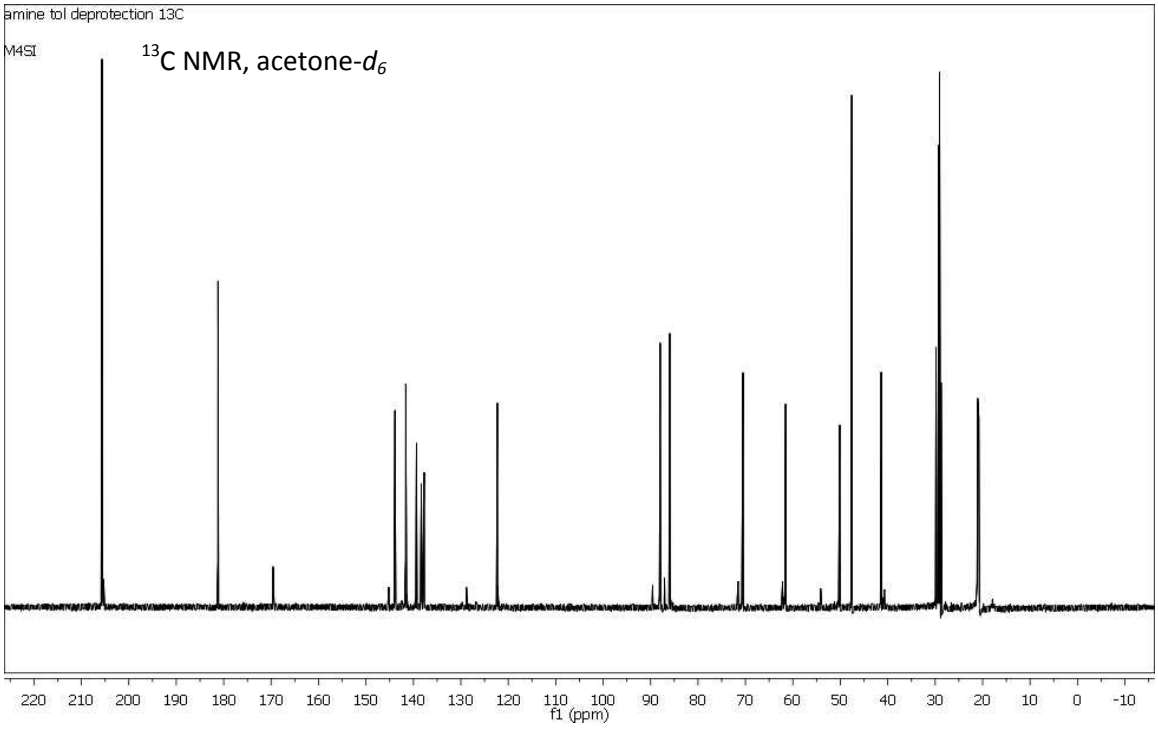

azide glycosylation publish 1H
^{1}H NMR, acetone-d_6
f1 (ppm)

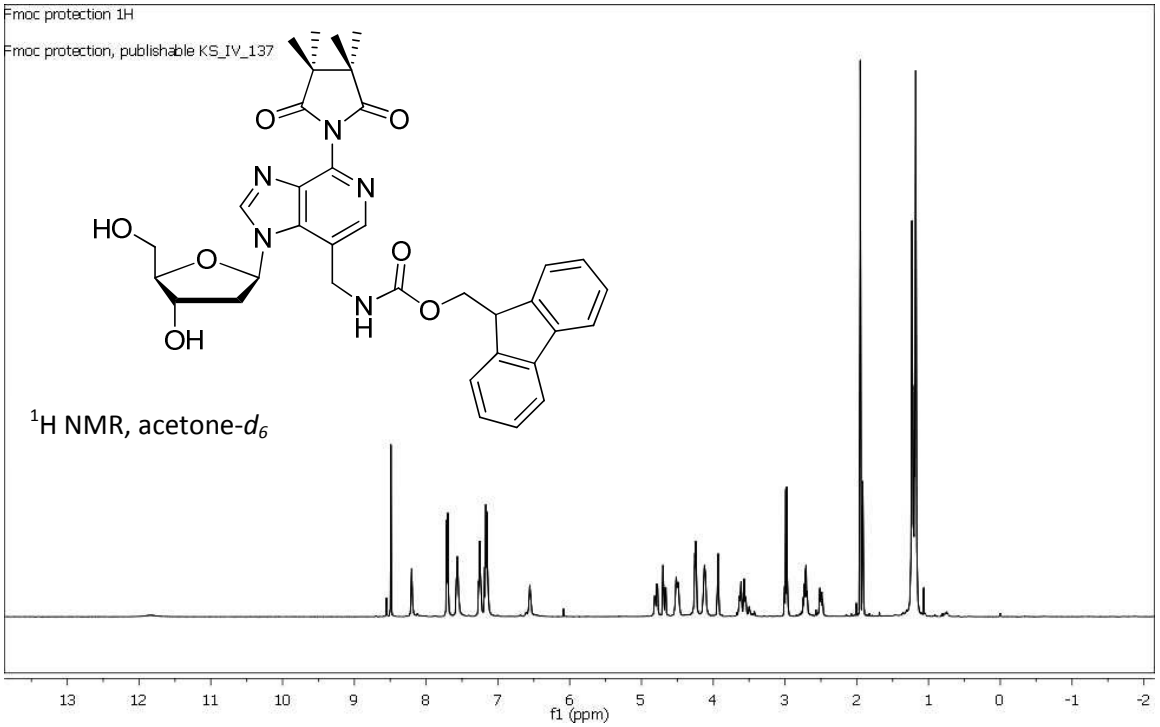

azide glycosylation publish 13C
aa
^{13}C NMR, acetone-d_6
f1 (ppm)

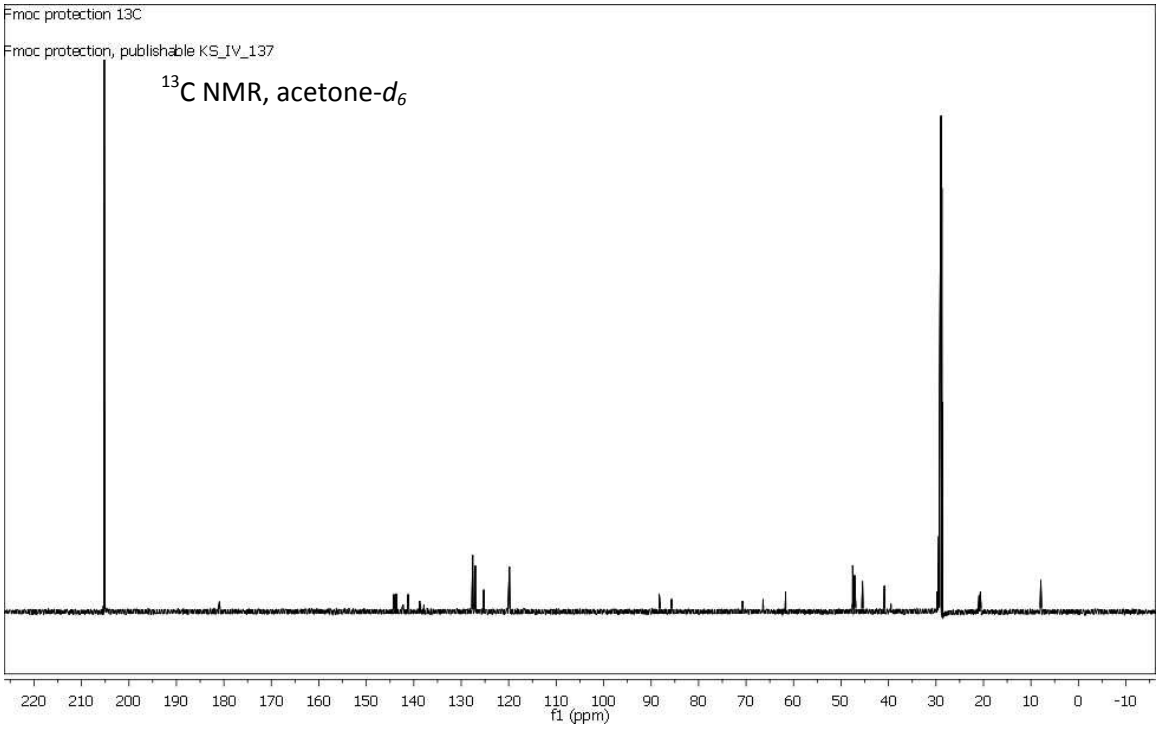

azide reduction publish 1H
^{1}H NMR, acetone-d_6
f1 (ppm)

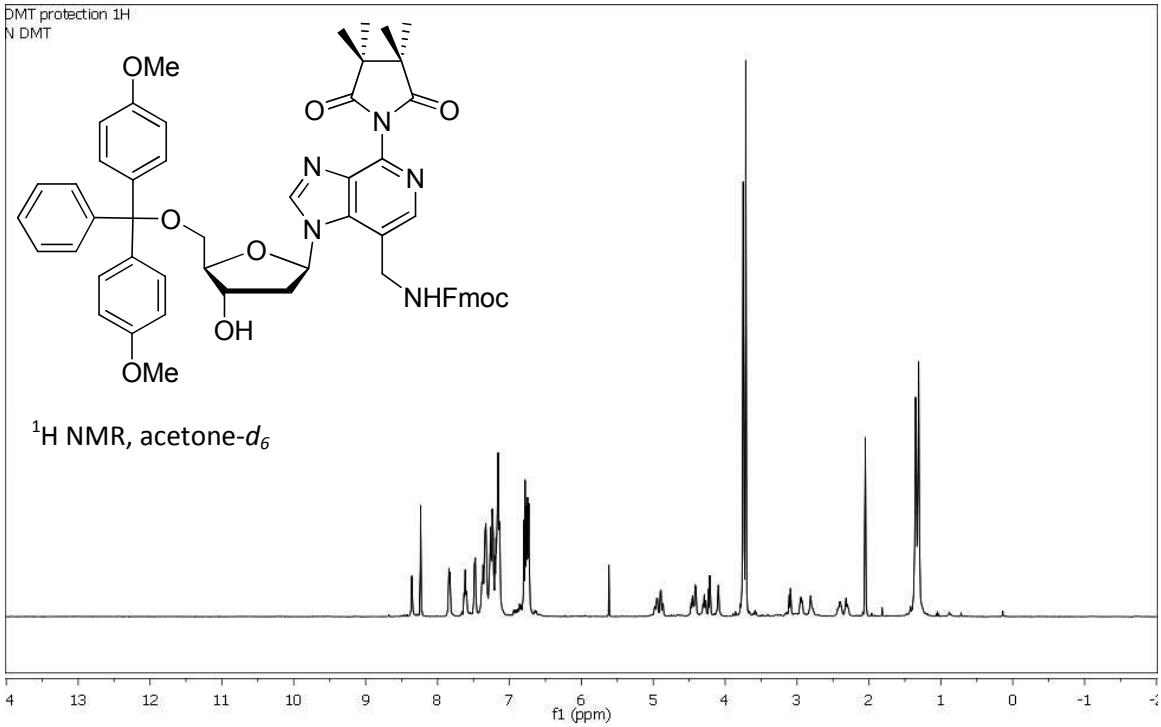

azide reduction publish 13C
aa
^{13}C NMR, acetone-d_6
f1 (ppm)

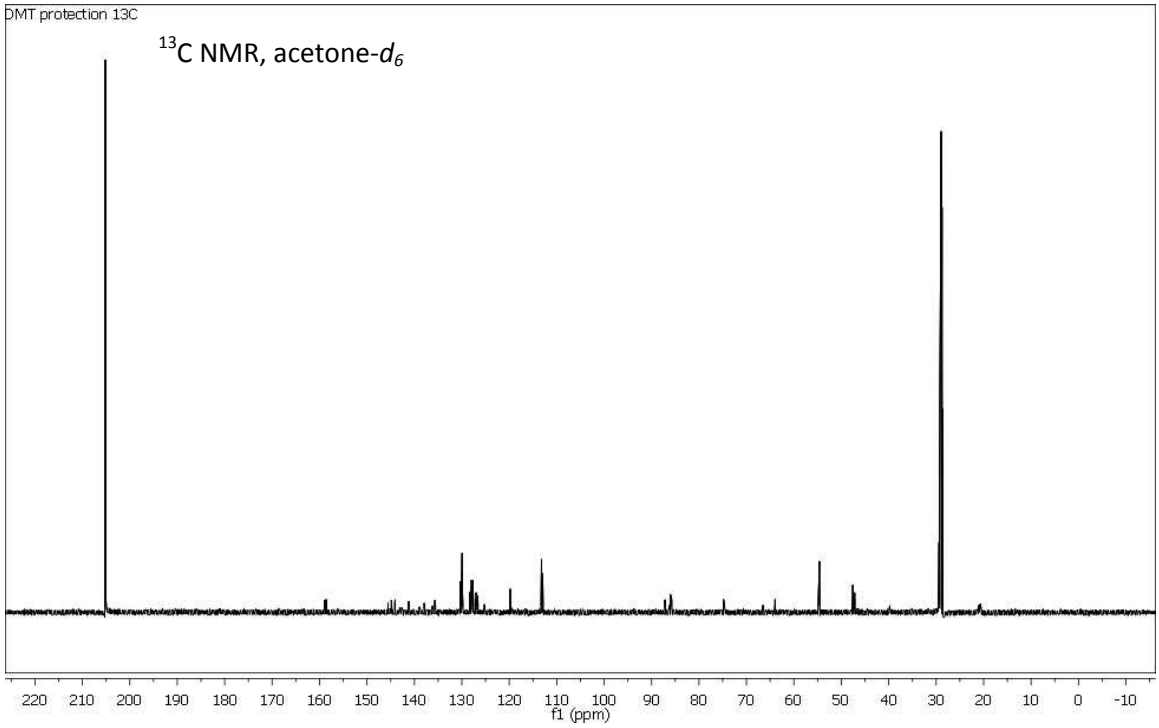

amine tol deprotection 1H
M4SI
HO
O
N
O
N
N
N
NH2
OH
1H NMR, acetone-d6
13 12 11 10 9 8 7 6 5 4 3 2 1 0 -1
f1 (ppm)

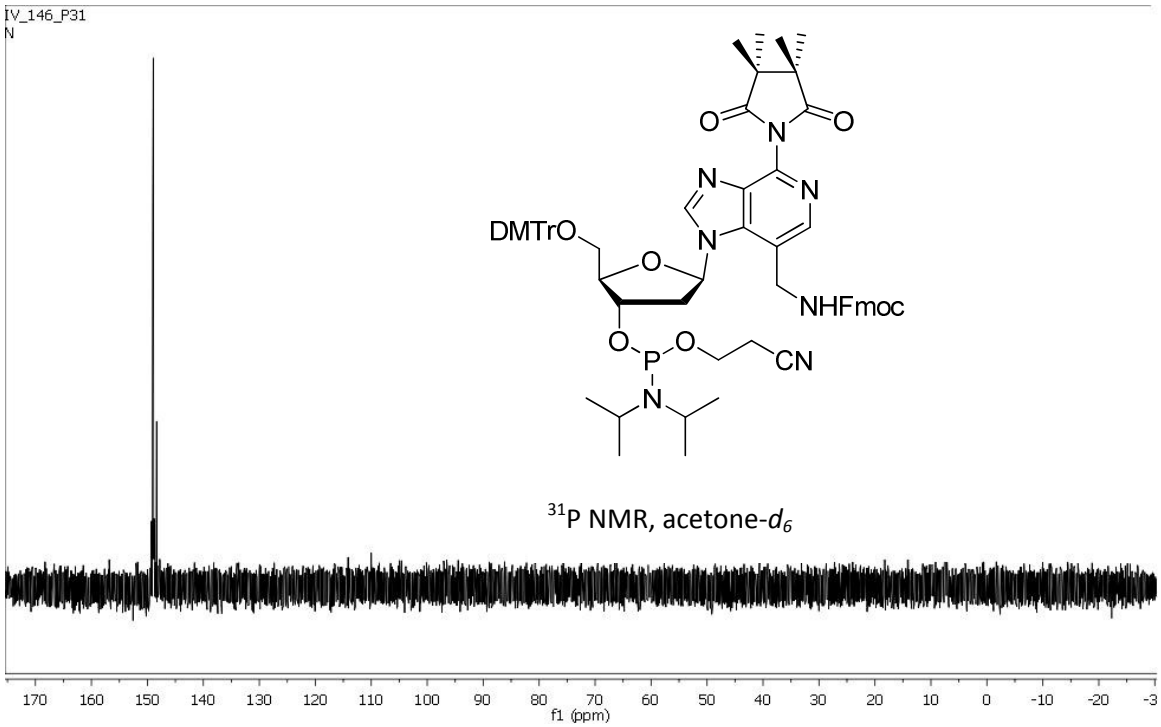

amine tol deprotection 13C
M4SI
13C NMR, acetone-d6
220 210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
f1 (ppm)


Fmoc protection 1H
Fmoc protection, publishable KS_IV_137
^{1}H NMR, acetone-d_6
f1 (ppm)

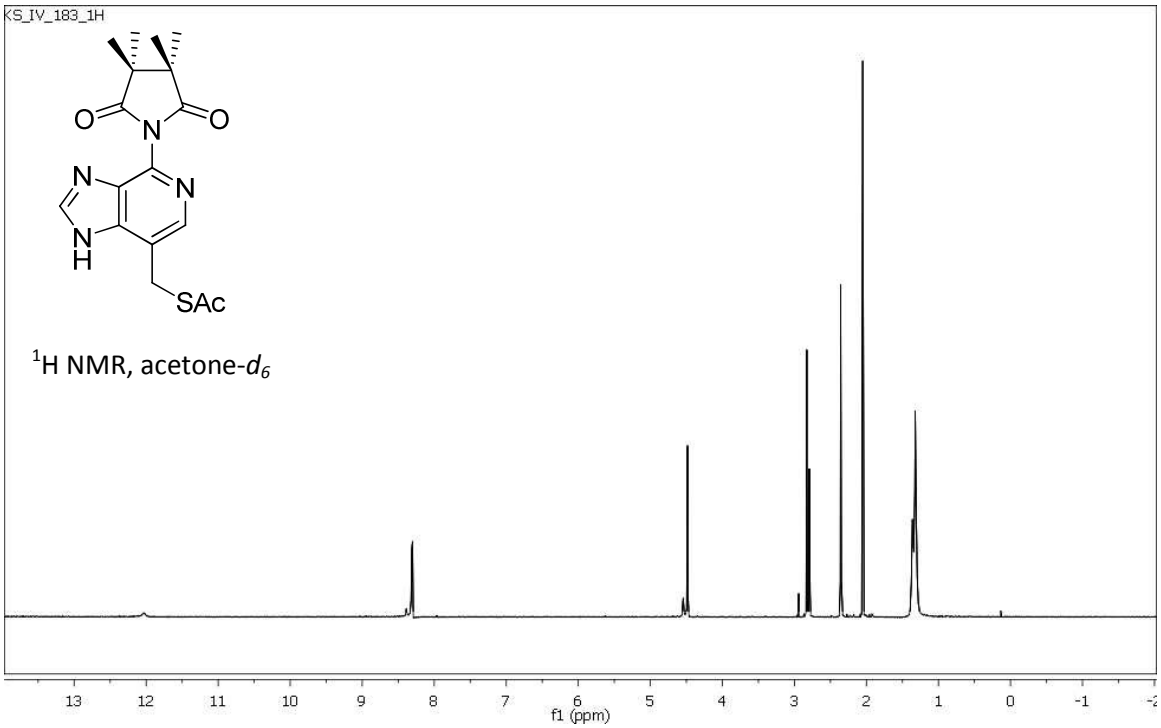

Fmoc protection 13C
Fmoc protection, publishable KS_IV_137
^{13}C NMR, acetone-d_6
f1 (ppm)


DMT protection 1H
N DMT
OMe
OMe
O
O
N
O
N
N
N
O
O
OH
NHFmoc
^{1}H NMR, acetone-d_6
f1 (ppm)

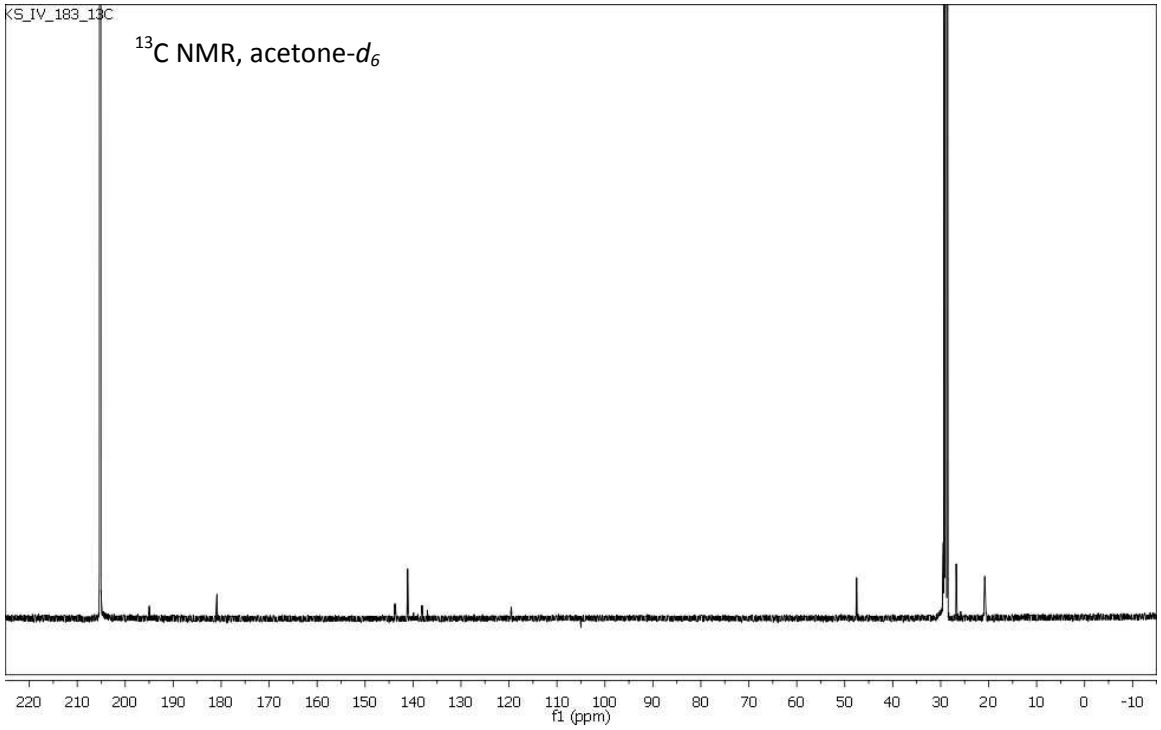

DMT protection 13C
^{13}C NMR, acetone-d_6
f1 (ppm)



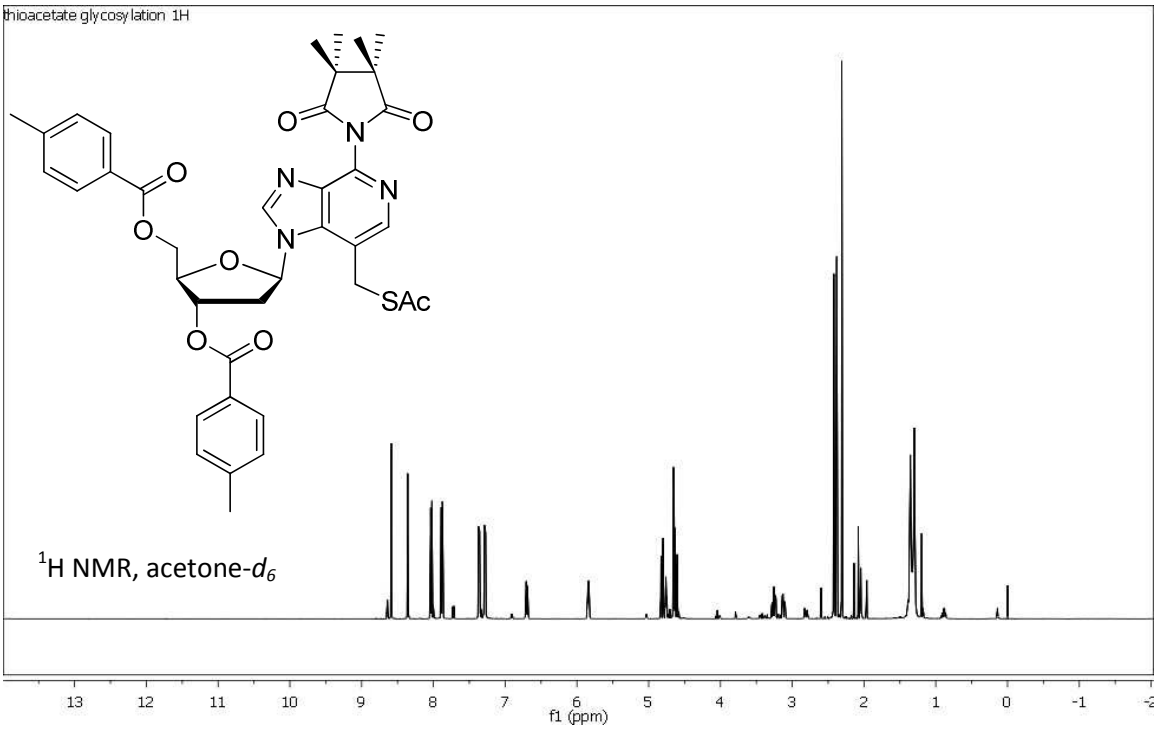
^{31}P NMR, acetone-d_6

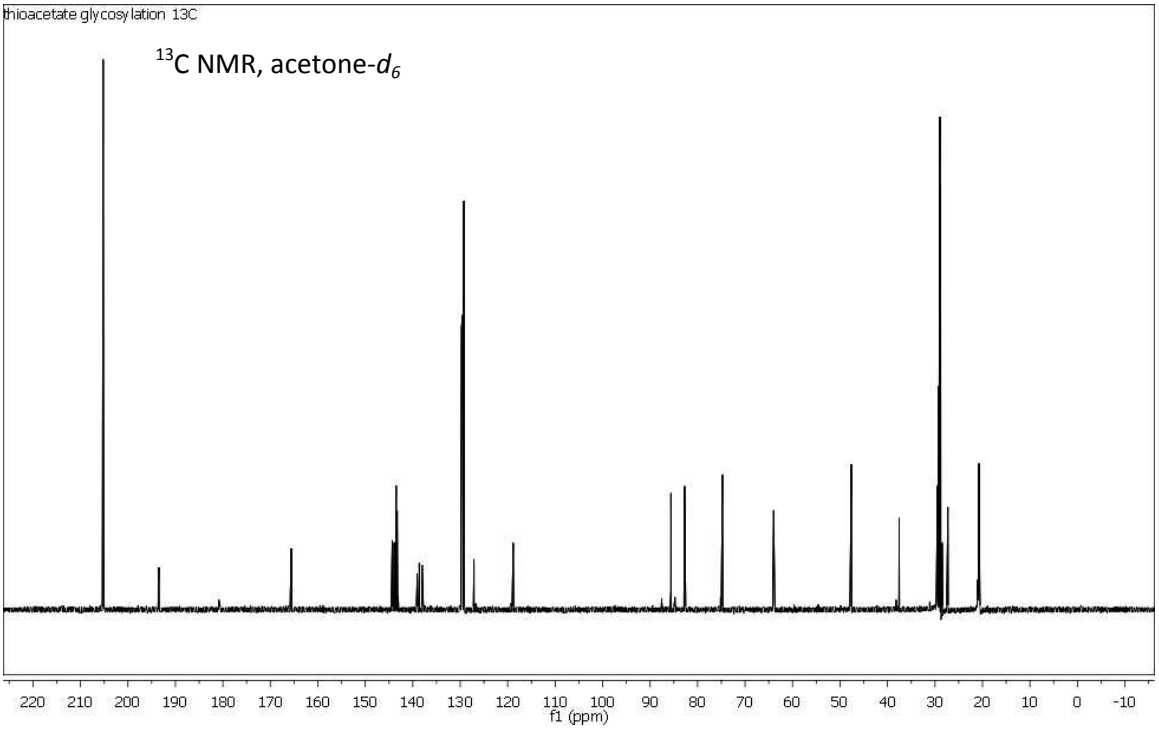


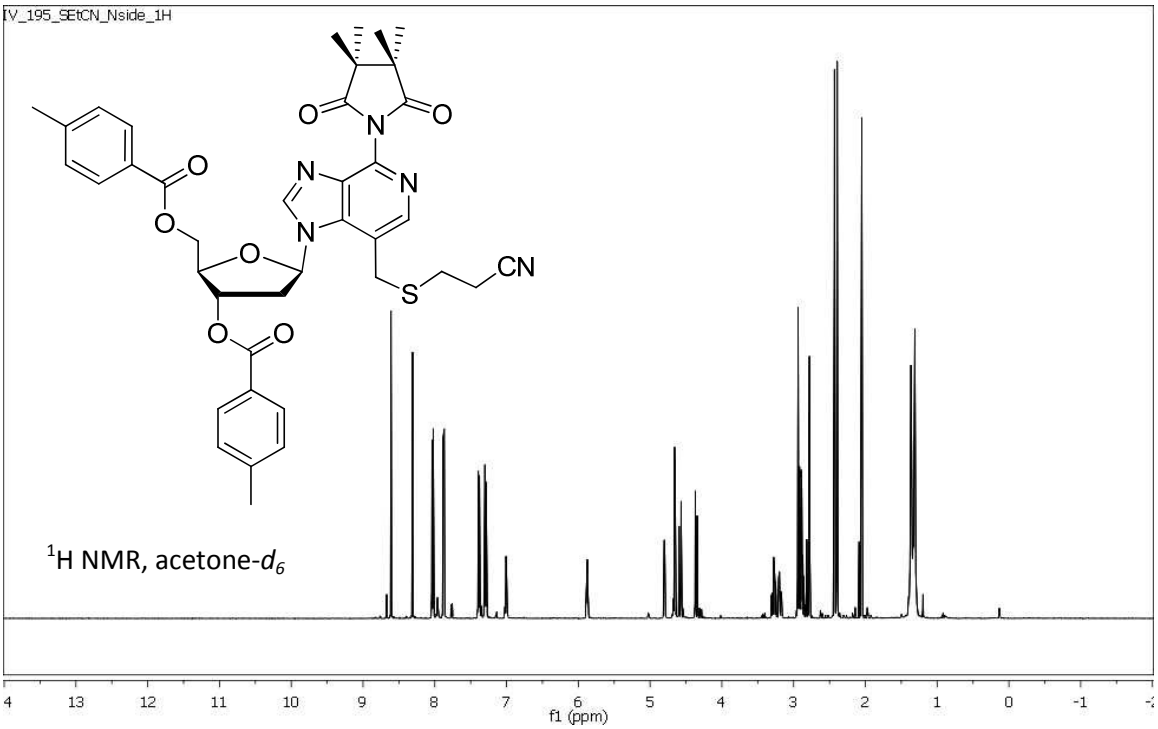
^{1}H NMR, acetone-d_6

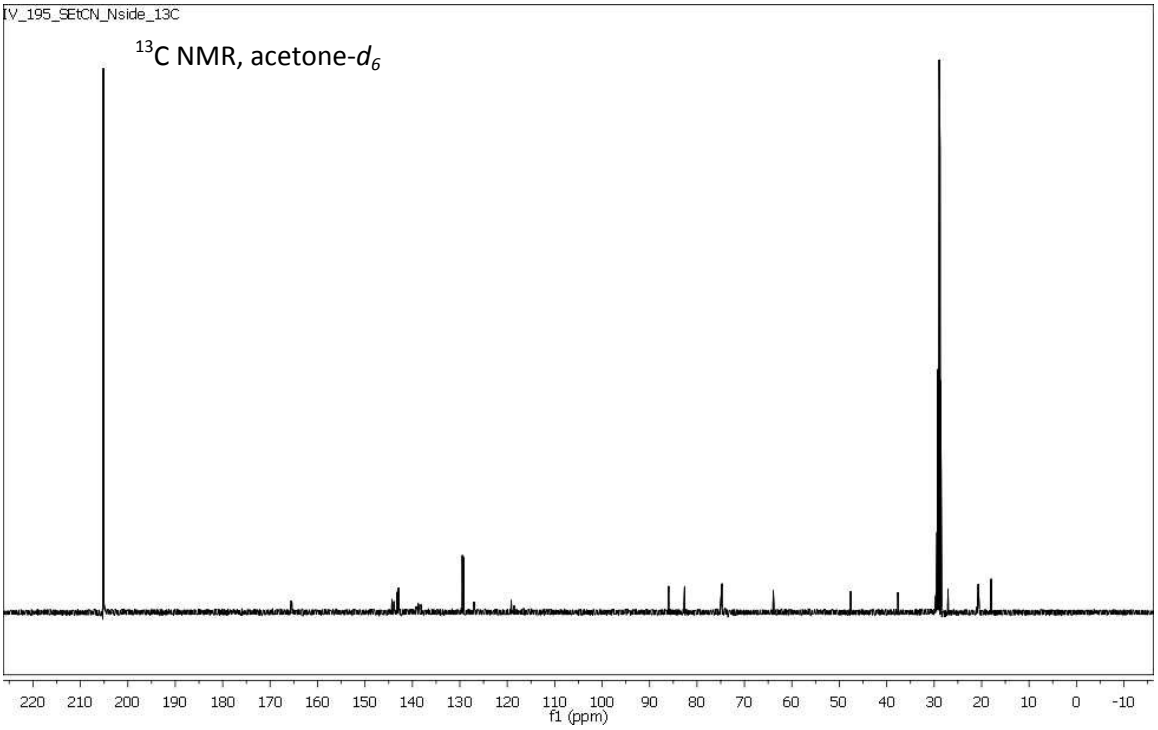


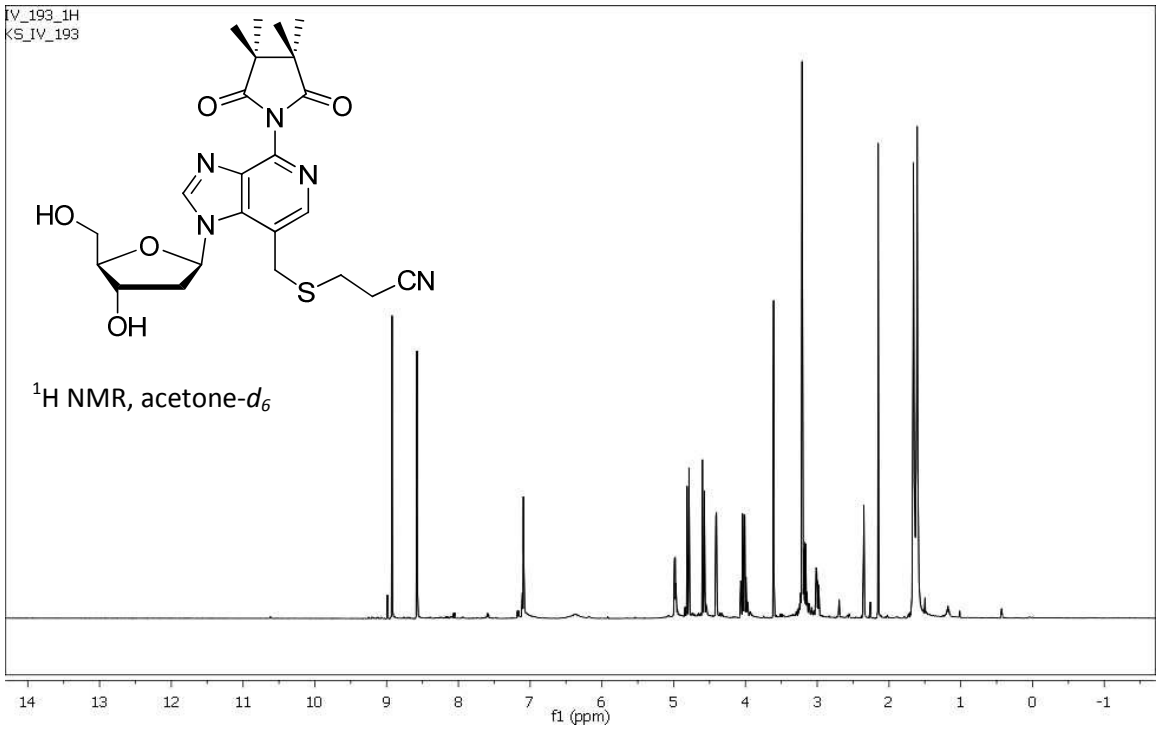
^{13}C NMR, acetone-d_6


thioacetate glycosylation 1H
SAc
^{1}H NMR, acetone-d_6
f1 (ppm)

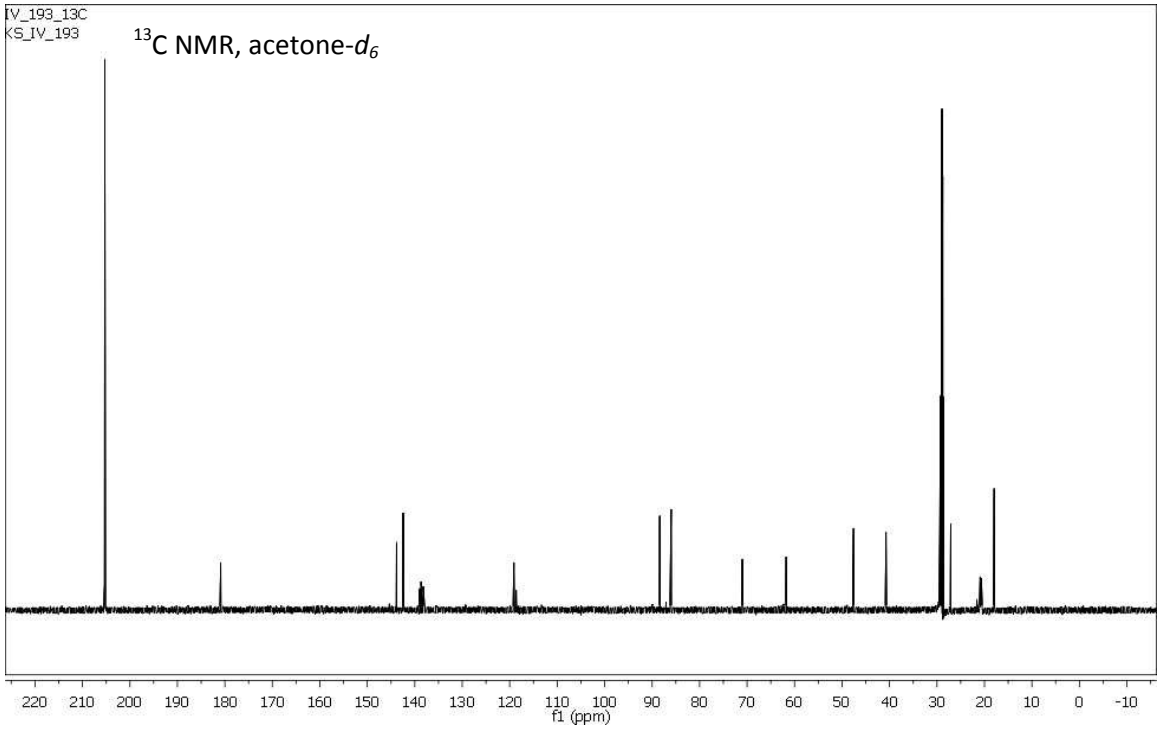

thioacetate glycosylation 13C
^{13}C NMR, acetone-d_6
f1 (ppm)


IV_195_SEtCN_Nside_1H
^{1}H NMR, acetone-d_6
f1 (ppm)

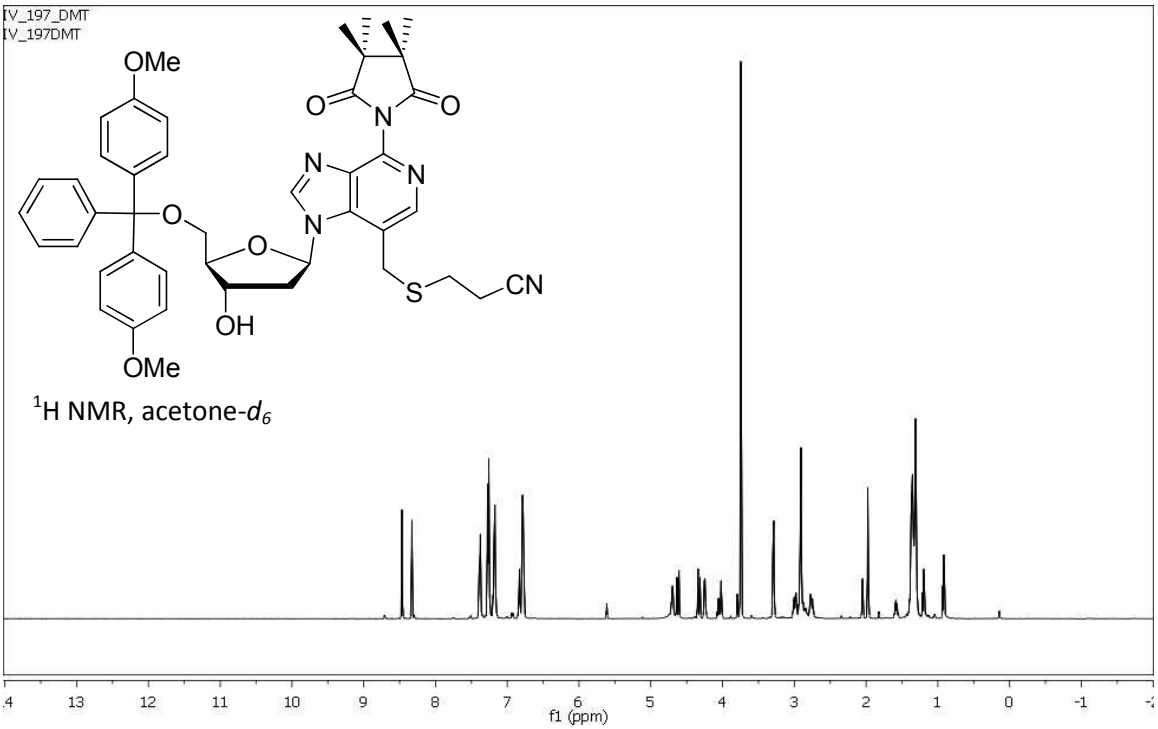

IV_195_SEtCN_Nside_13C
^{13}C NMR, acetone-d_6
f1 (ppm)

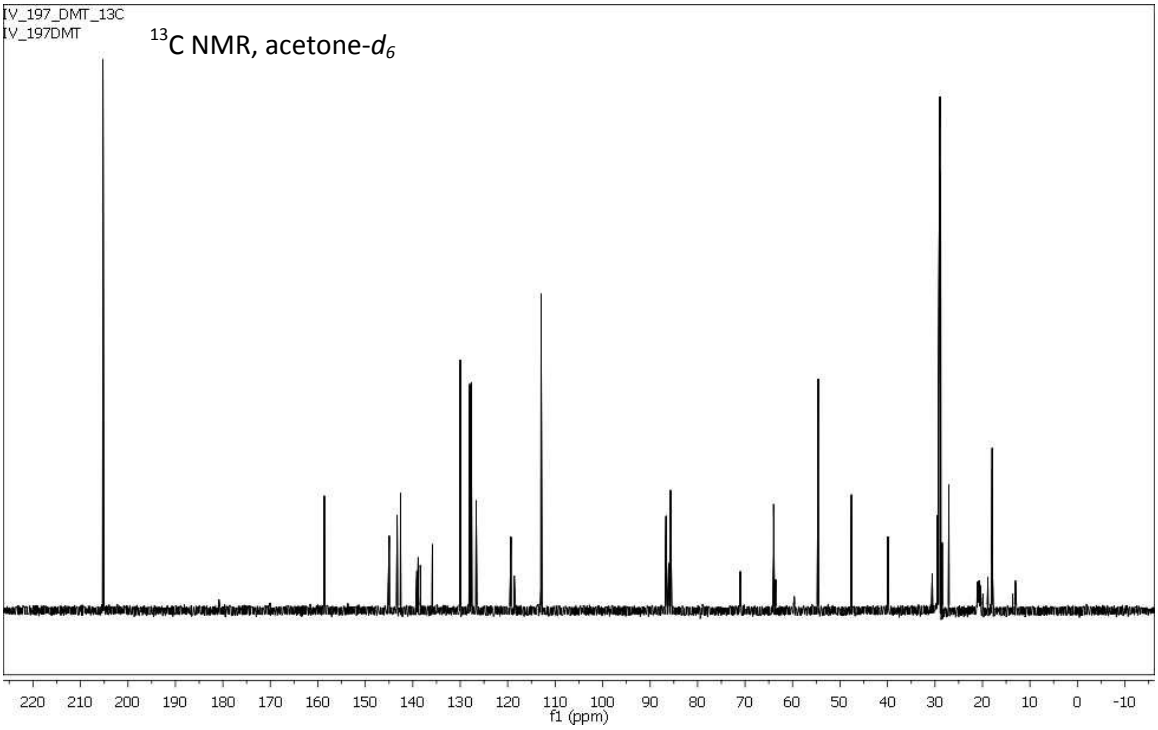


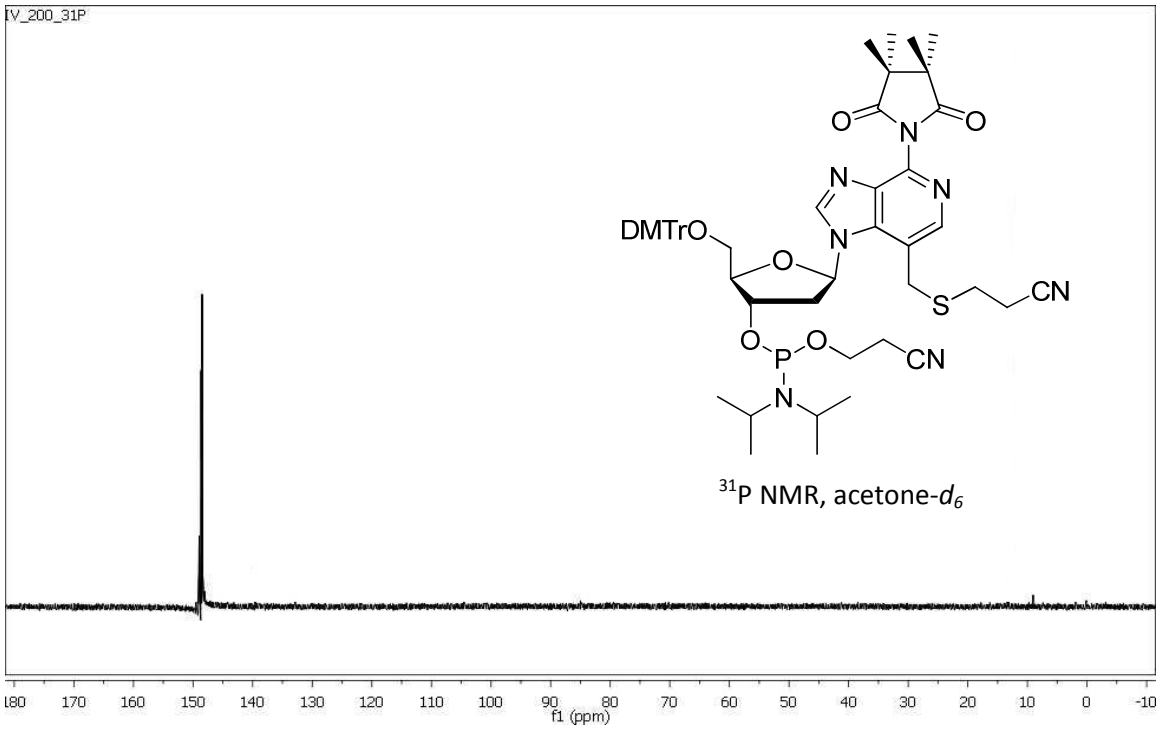
^{1}H NMR, acetone-d_6



^{13}C NMR, acetone-d_6


IV_197_DMT
IV_197DMT
OMe
OMe
O
O
N
O
N
N
N
O
OH
S
CN
^{1}H NMR, acetone-d_6
f1 (ppm)


IV_197_DMT_13C
IV_197DMT
^{13}C NMR, acetone-d_6
f1 (ppm)



^{31}P NMR, acetone-d_6

Appendix II: α-Plots and van Hoft Analysis of Oligonucleotides

Thermal denaturation studies were performed in solutions of 20 mM NaH_2PO_4 (pH 7.0) and 1 M NaCl with the concentrations of duplex at 5 μM). Absorbance and temperature values were measured with an AVIV 14DS UV-visible spectrophotometer equipped with digital temperature control. The temperature of the cell compartment was increased in 1.0 °C/min steps (from 4 to 95 °C), and when equilibrium was reached, temperature and absorbance data were collected every 1 °C. The raw data was exported to the software application Microcal Origin (OriginLab Corporation, Northampton, MA) for determination of the melting temperature (T_m) and free energy (ΔG) of each modified duplex. First, the T_m plot (absorbance vs temperature) was converted to an alpha (α) plot (fraction of molecules paired vs temperature) using the equation:

$$\boldsymbol{\alpha} = \frac{\boldsymbol{A - A_s}}{\boldsymbol{A_d - A_s}} \quad (1)$$

where A is the absorbance at a given temperature, and A_s and A_d are the absorbance values for the single- and double-stranded states, respectively. The parameter α represents the fraction of molecules paired; the T_m is extrapolated when $\alpha = 0.5$. In order to obtain sufficient approximations for A_s and A_d, the upper and lower baselines of the melting curve were fitted using linear least-squares fits. The thermodynamic parameters ΔH, ΔS, and ΔG were determined using a van't Hoff plot (ln K vs $1/T$). The value of K was calculated at each temperature from the α plot using the following equation:

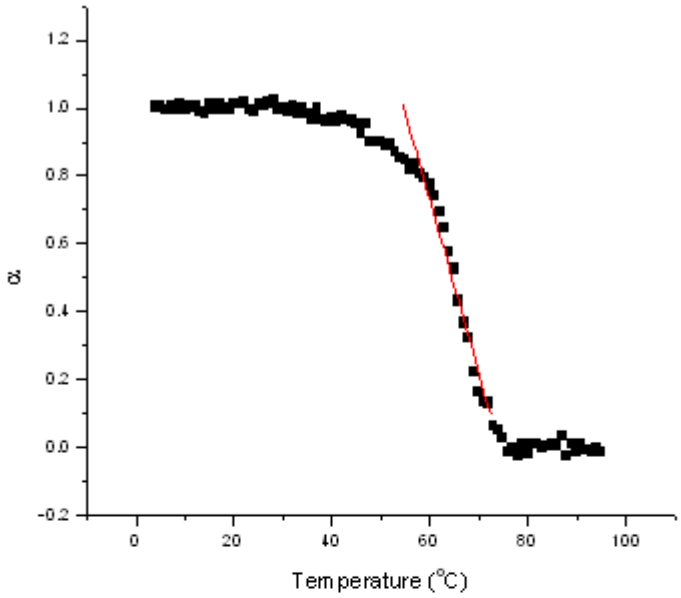
$$K = \frac{\boldsymbol{\alpha}}{\boldsymbol{2(1-\alpha)^2 C_t}} \quad (2)$$

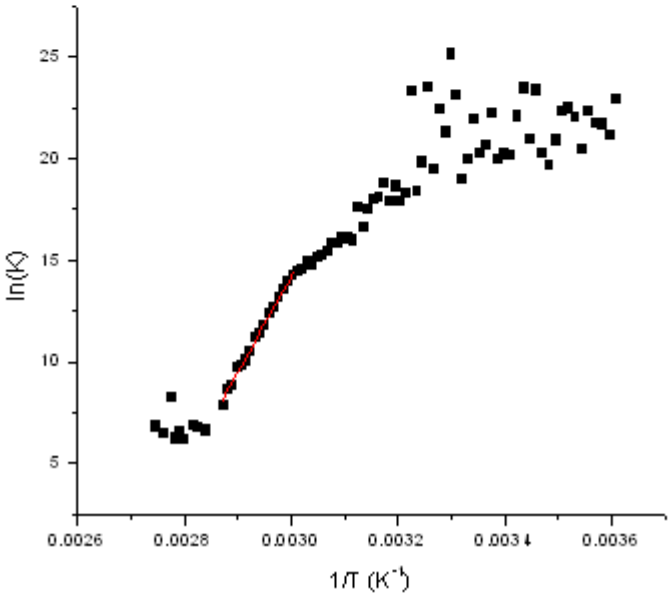
where C_t is the total strand concentration. The thermodynamic parameters can be derived directly from the linear least-squares fit in which the slope is $-\Delta H^o/R$ and the intercept is

$\Delta S^o/R$. The free energy change (ΔG^o) at any temperature can be determined by using the following relationship:

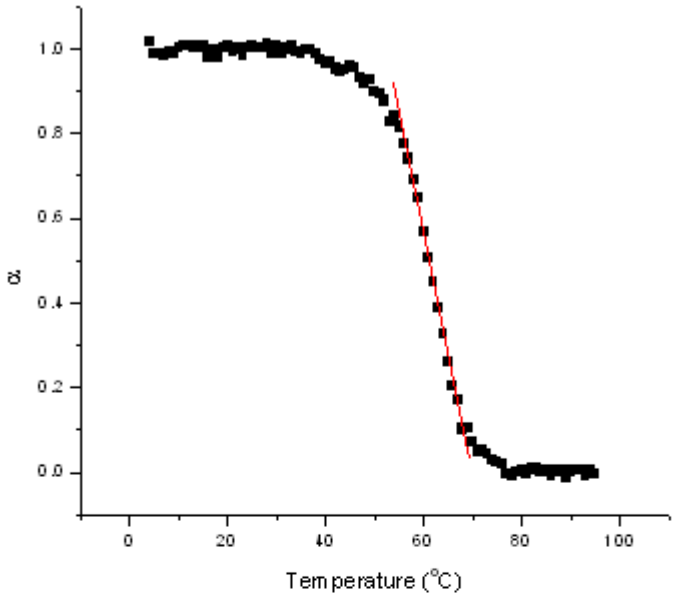
$$\Delta G° = \Delta H° - T\Delta S° \quad (3)$$

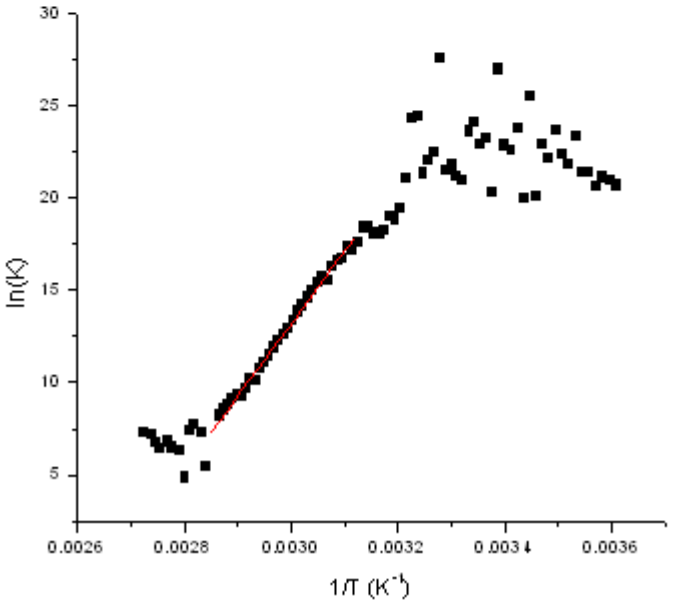
d(CCGG AAAA CGCC)



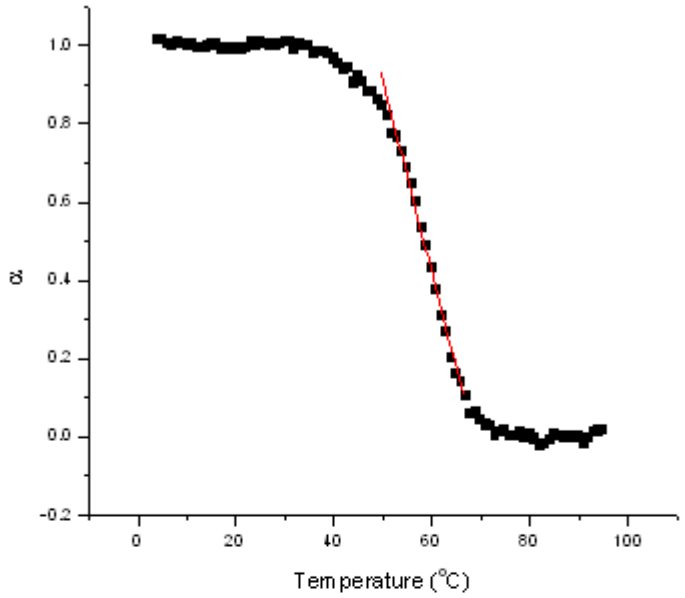


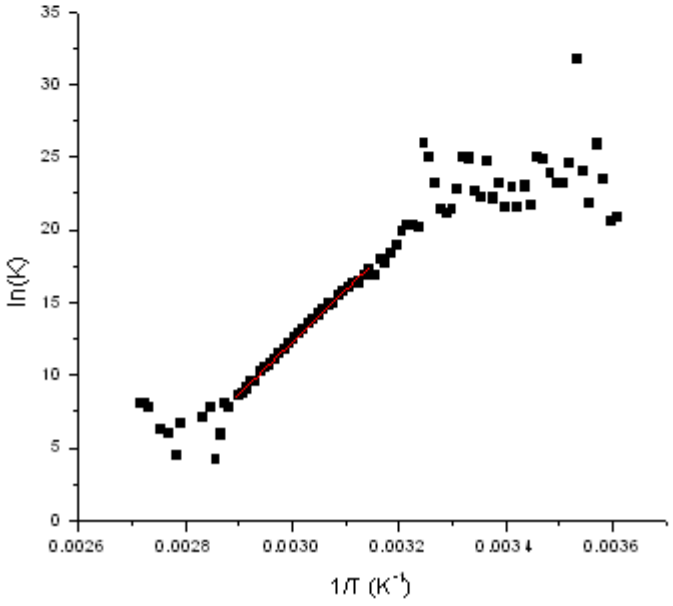
d(CCGG AA$\boldsymbol{A}^{\boldsymbol{H}}$A CGCC)



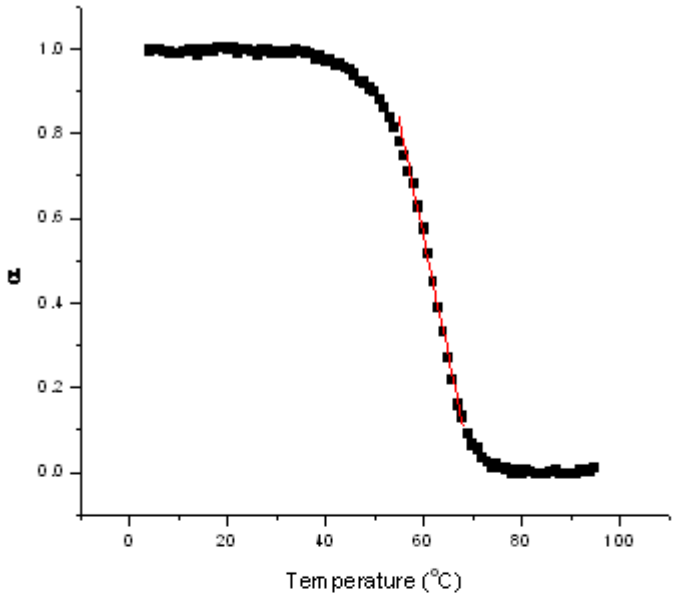


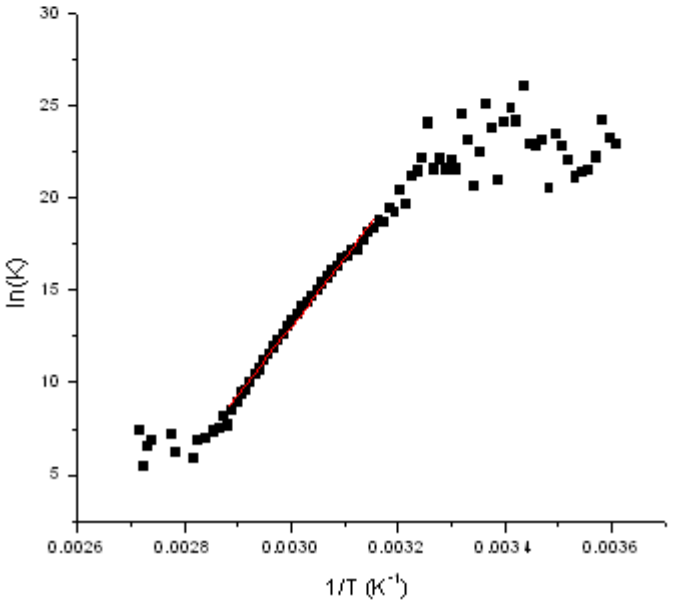
d(CCGG AA$\boldsymbol{A}^{CH3}$A CGCC)



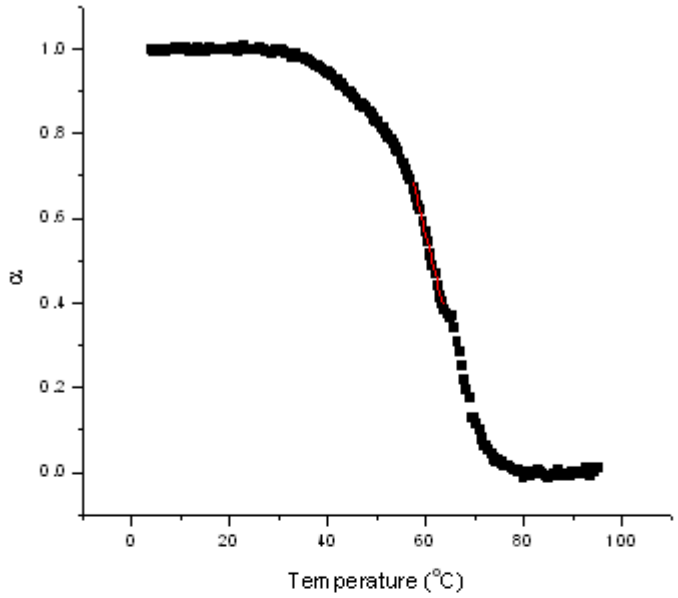


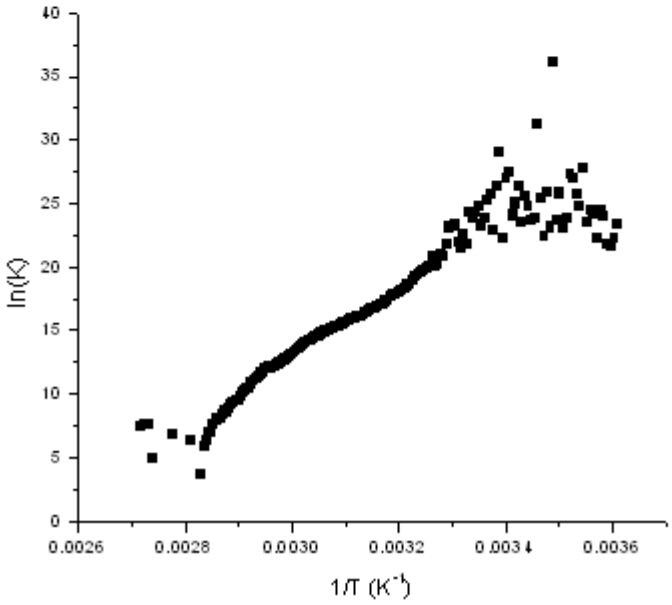
d(CCGG AA***A***^{CH2OH}A CGCC)



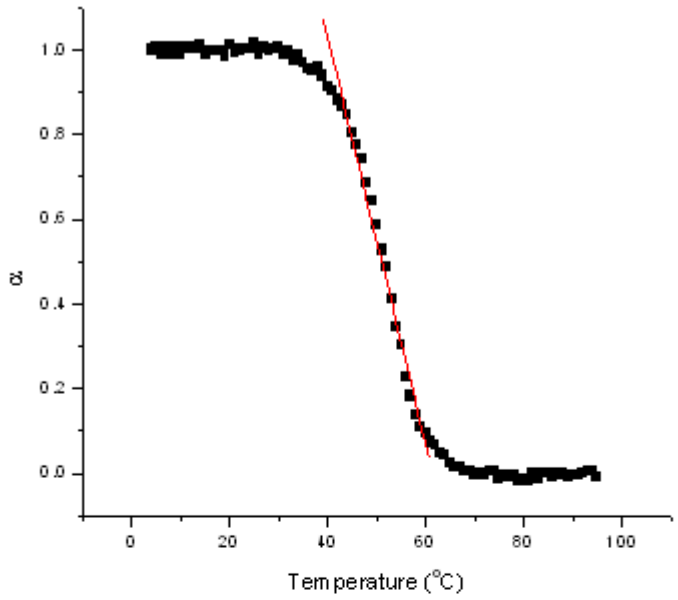


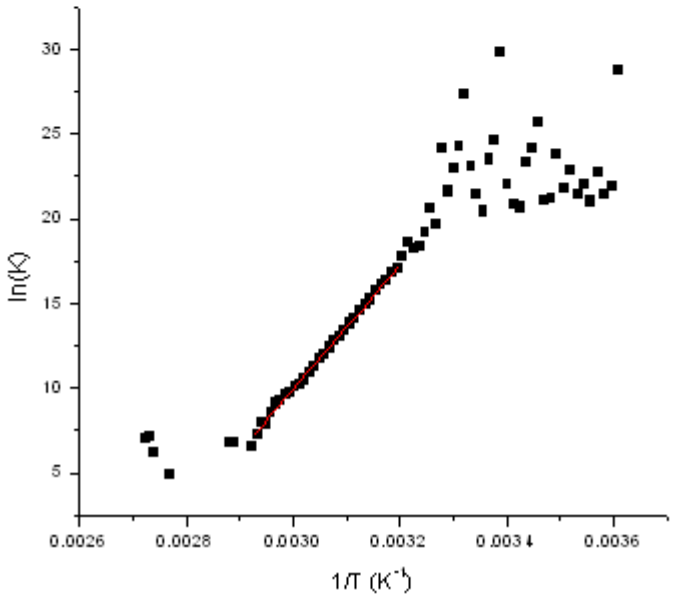
d(CCGG AA***A***$^{CH2NH3+}$A CGCC)



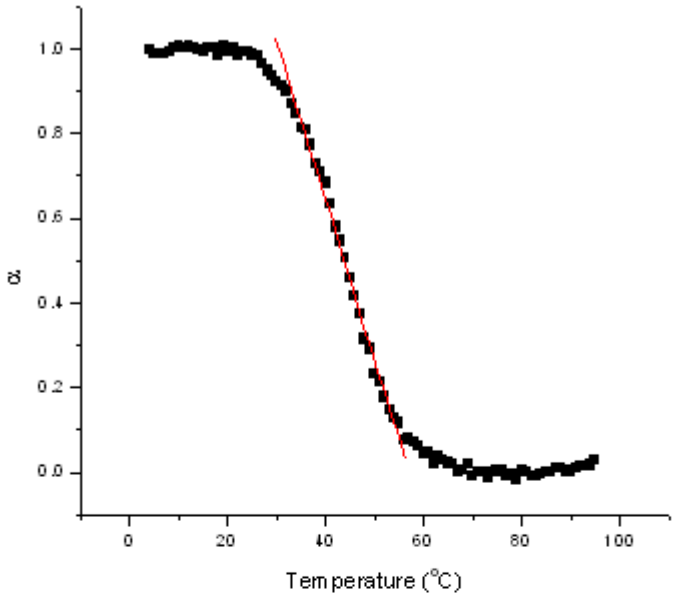


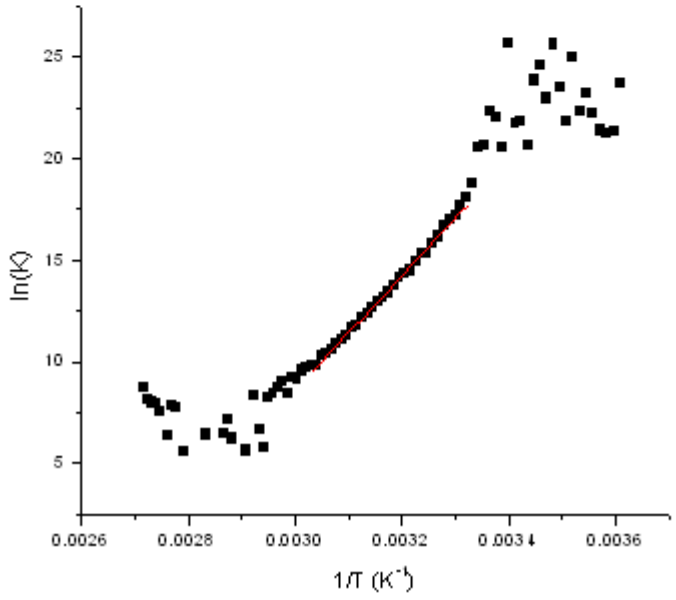
d(CCGG ($\boldsymbol{A}^{\boldsymbol{H}}$)$_4$ CGCC)



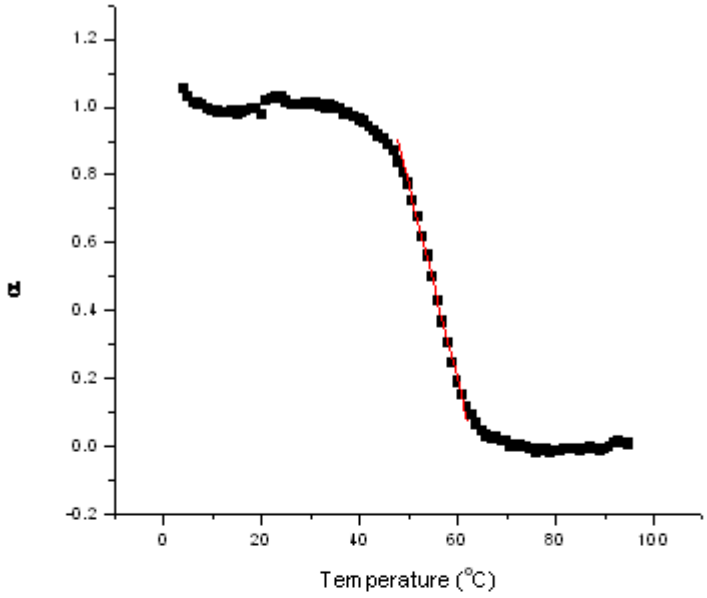


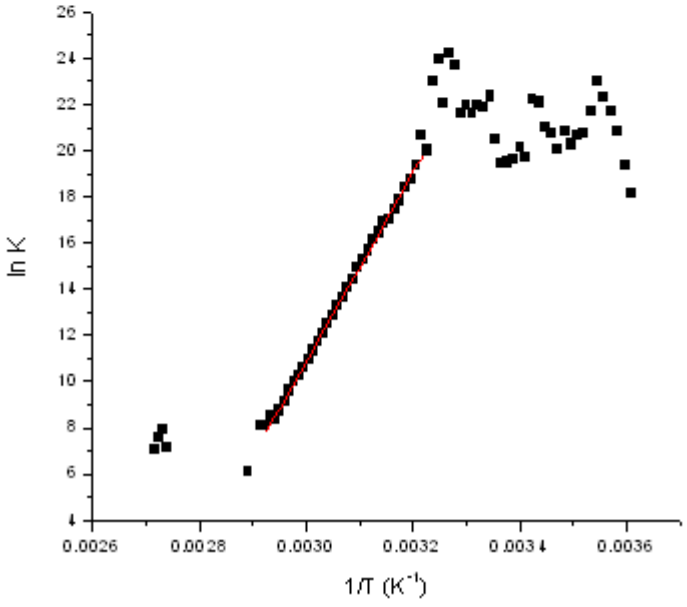
d(CCGG (A^{CH3})$_4$ CGCC)



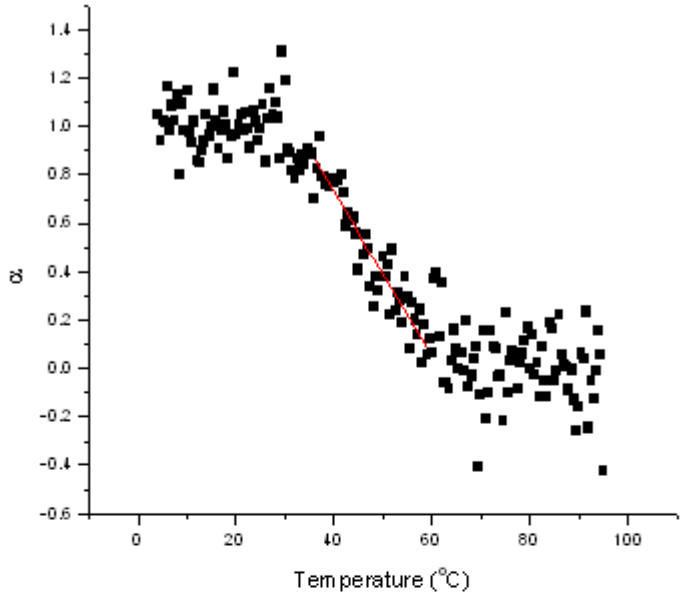


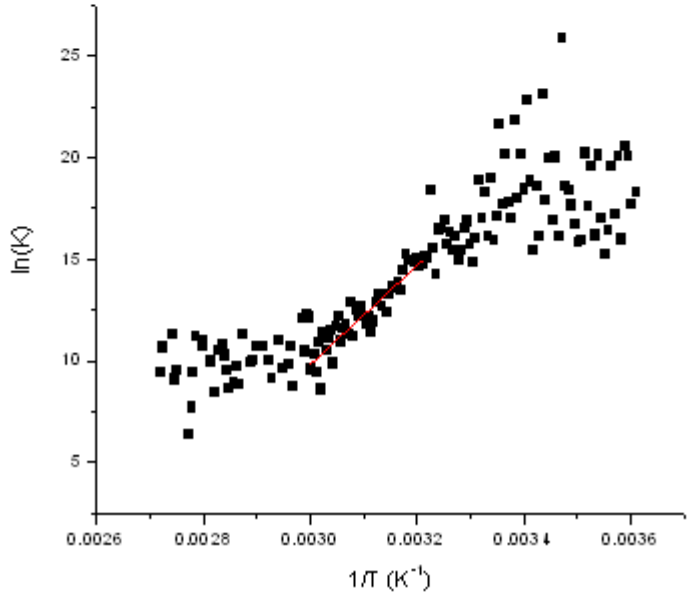
d(CCGG (A^{CH2OH})$_4$ CGCC)



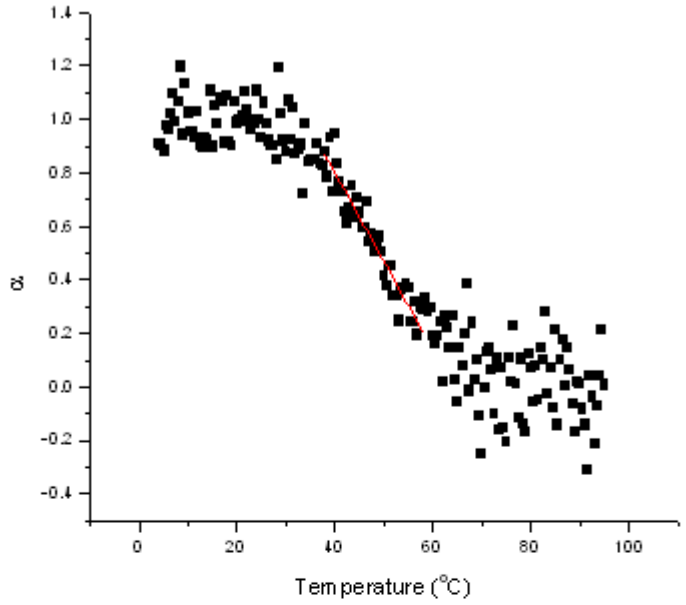


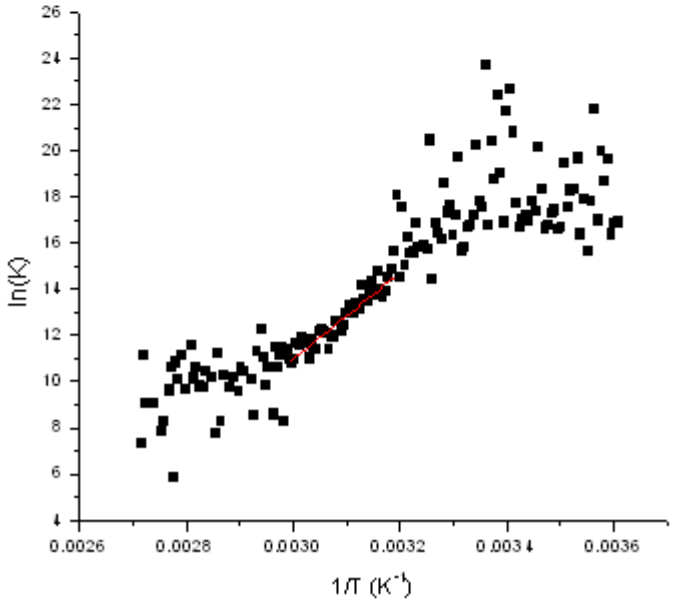
d(CCGG ($A^{CH2NH3+}$)$_4$ CGCC)



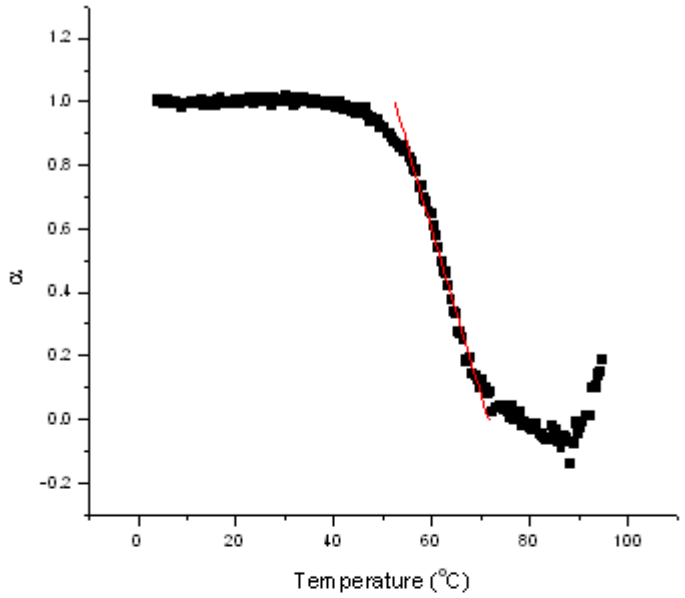


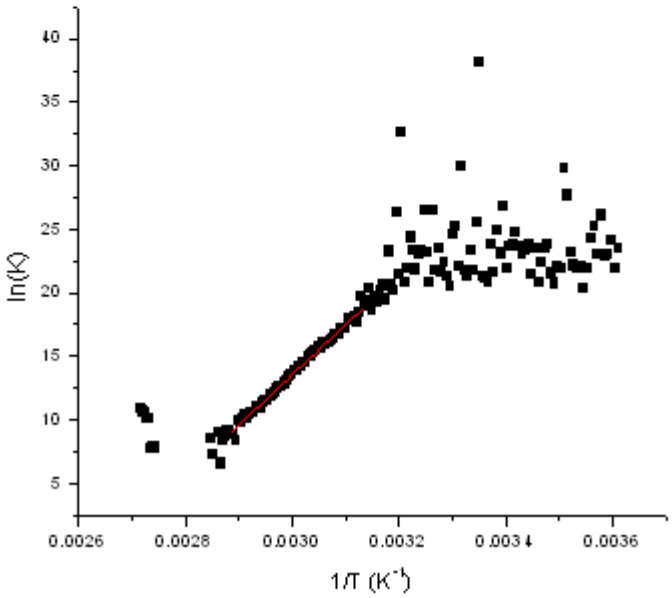
d(CCGG (A^{CH2NH2})$_4$ CGCC), pH 9



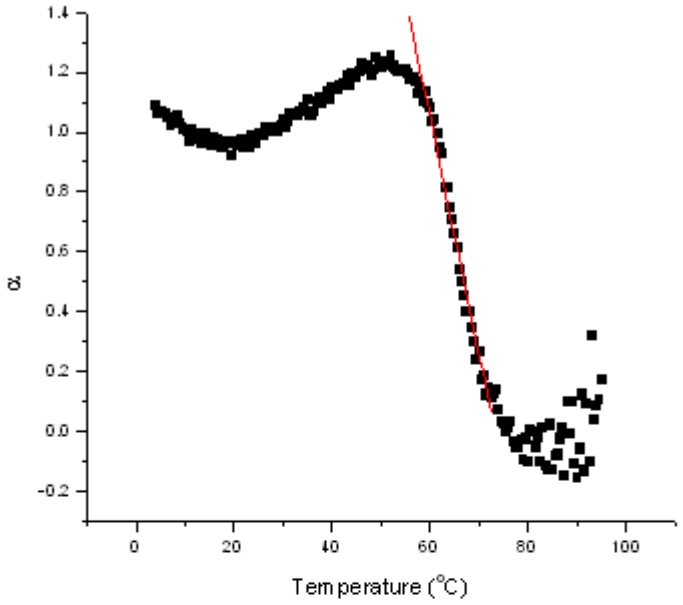


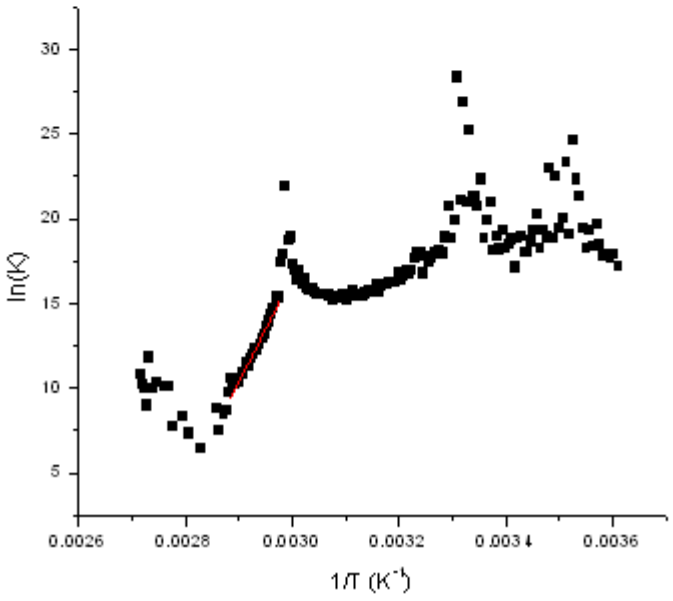
d(CCGG $\boldsymbol{A}^{CH2NH3+}$A$\boldsymbol{A}^{CH2NH3+}$A CGCC)





d(CCGG AA***A***^{CH2NHR}A CGCC) with Marina Blue





d(CCGG AA***A***^{CH2NHR}A CGCC) with Fluorescein-5-EX

It should be noted that the presence of the fluorescein-5-EX group severely distorts the sigmoidal shape of the oligo's melting curve. Methods for calculating α- plots and van Hoft plots are inapplicable to this type of curve. The melting temperature is extrapolated from the positively sloped line.